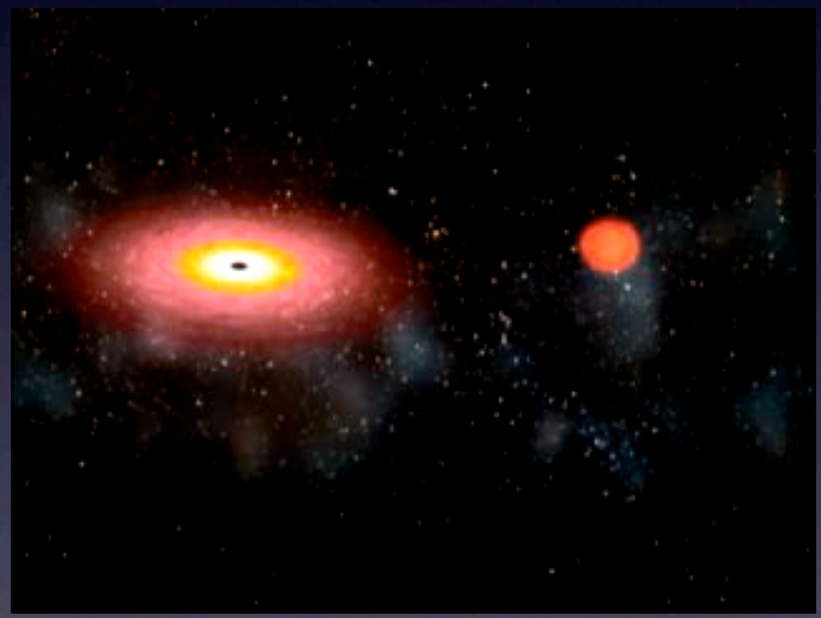
LIGO Global Computing in the Next Decade

Patrick Brady for the LIGO Scientific Collaboration
University of Wisconsin-Milwaukee

Compact Binaries

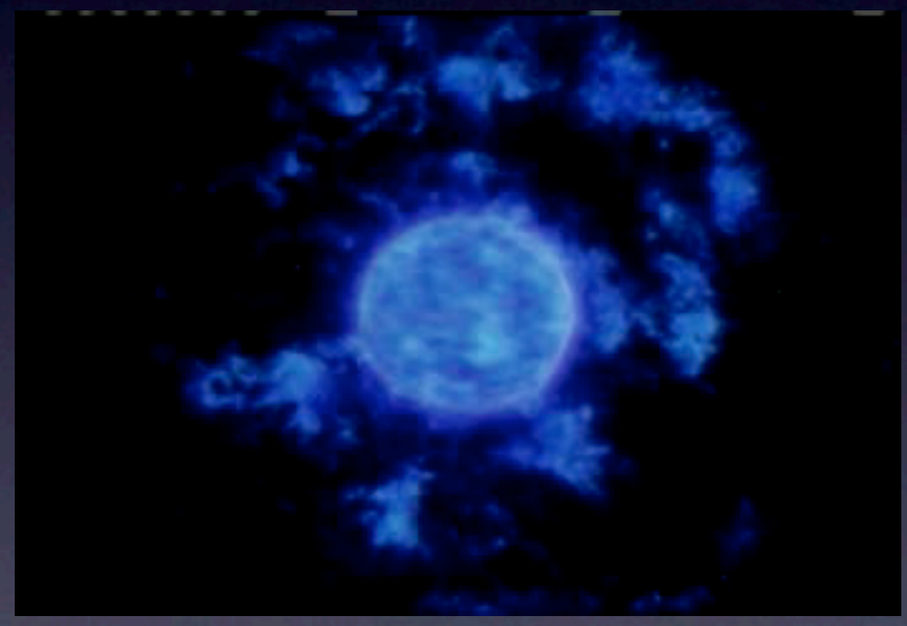
- Pairs of black holes, neutron stars, or a black hole and neutron star
- As they orbit one another, they emit gravitational waves and the objects get closer together, eventually merging
- LIGO is sensitive to last few minutes of inspiral and merger



Gravitational-wave bursts

- Very short (< 1 sec) bursts of waves from violent astrophysical events

- Examples include supernova explosions, mergers of compact binaries, and cosmic string kinks and cusps

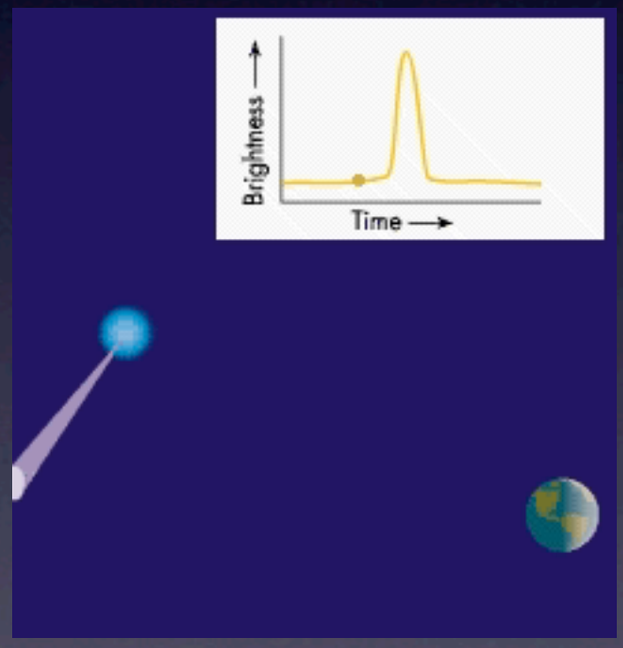


Continuous Signals

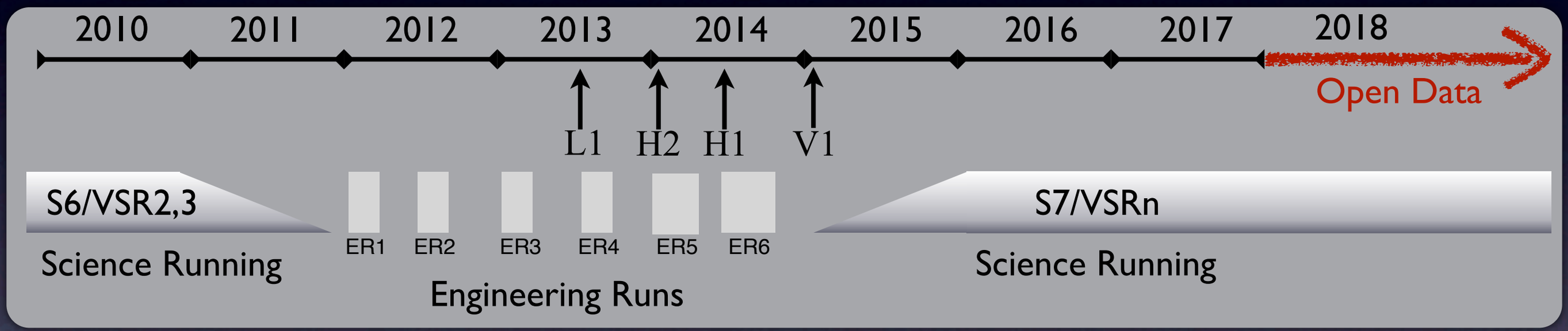
- Signals lasting as long as, or longer than, the observation time

- Known radio pulsars could also emit gravitational waves

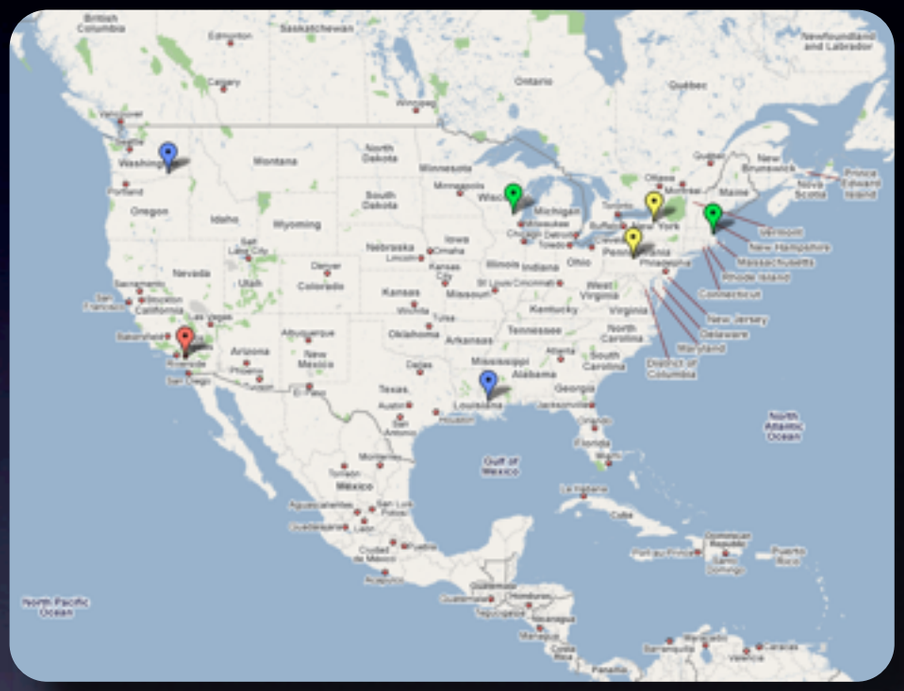
- Unknown radio pulsars that are not beamed toward earth

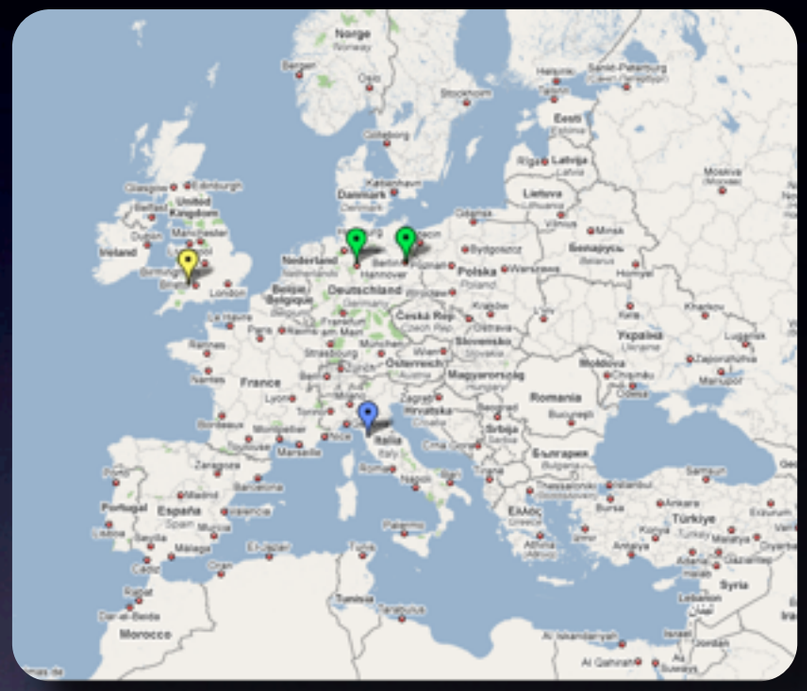


From LIGO to aLIGO



Global Network of Gravitational-wave Detectors





Observatories

Tier-1 Center

Tier-2 Center

Tier-3 Center

- Observatories: LIGO (2 US) and Virgo (1 EU)
- Future: LCGT (Japan) and LIGO (Australia)
- LIGO Data Grid (LDG): combination of computational and data storage resources with grid-computing middleware to create a distributed gravitational-wave data analysis facility.

Current Capabilities

- 5 US & 3 EU centers
  - ~20,000 CPU cores
  - Cluster level scheduling
  - Reference operating system
- LIGO Data Replicator
  - Bulk data replication to sites
- MetaDatabases
  - Federated, web & command-line user interfaces
- Software bundle
  - Packaged for reference OS and distributed using native packaging

Functional system for offline analysis

aLIGO Data Products

- Time series in aLIGO
  - Raw, calibrated, cleaned, noise, data quality. 600 TB per year; 1.8 PB allowing for 300% copies
- Currently maintain 1.8PB/42M files of LIGO data in central archive at Caltech with redundant copies at the Observatories.
- Actively manage & regularly back up ~100TB/ 600M files of user data products.
- Regularly distribute 10-100TB datasets world wide to Tier-2 center local disk caches.
- Must bring all data together in one place for joint analysis.

Observatories: write raw data, generate Reduced Data Sets, catalog files, write one copy to **disk**, LIGO Data Replicator to Central Archive & **2nd site**

↓

Central Archive: **3rd disk & tape** copy, **all** data cached to disk, RDS & h(t) to Tier 2

↓

Tier 2 Centers: Locally cache **all RDS & h(t)** data

aLIGO Data Products
(no LIGO equivalent)

- Catalogs
  - Data quality: small database of times at present, could grow to include better characterization on noise transients, etc
  - Objects, measured parameters: requirements are uncertain at this time. Even in routine detection era for aLIGO should be easily handled.
  - Plan regular releases once science running begins
- Transient (compact binaries & bursts) alerts
  - Outgoing alerts with low-latency (seconds after data acquisition) to allow follow-up with other instruments
  - Incoming for rapid gravitational-wave follow-up

Gravitational-wave Astronomy

- Robust online and offline data handling and analysis capabilities are required
- Online system:
  - Pipelines generating transient alerts & data quality information within seconds of data acquisition
  - Careful attention to interfaces between control/diagnostic systems, data acquisition systems, and processing systems through close collaboration
- Offline system:
  - Re-processing of data for transients including deeper searches
  - Enhanced data quality generation
  - Searches for continuous and stochastic signals
  - Parameter estimation, model selection and simulations

aLIGO Computational Requirements Scaled from LIGO analysis & Moore's Law

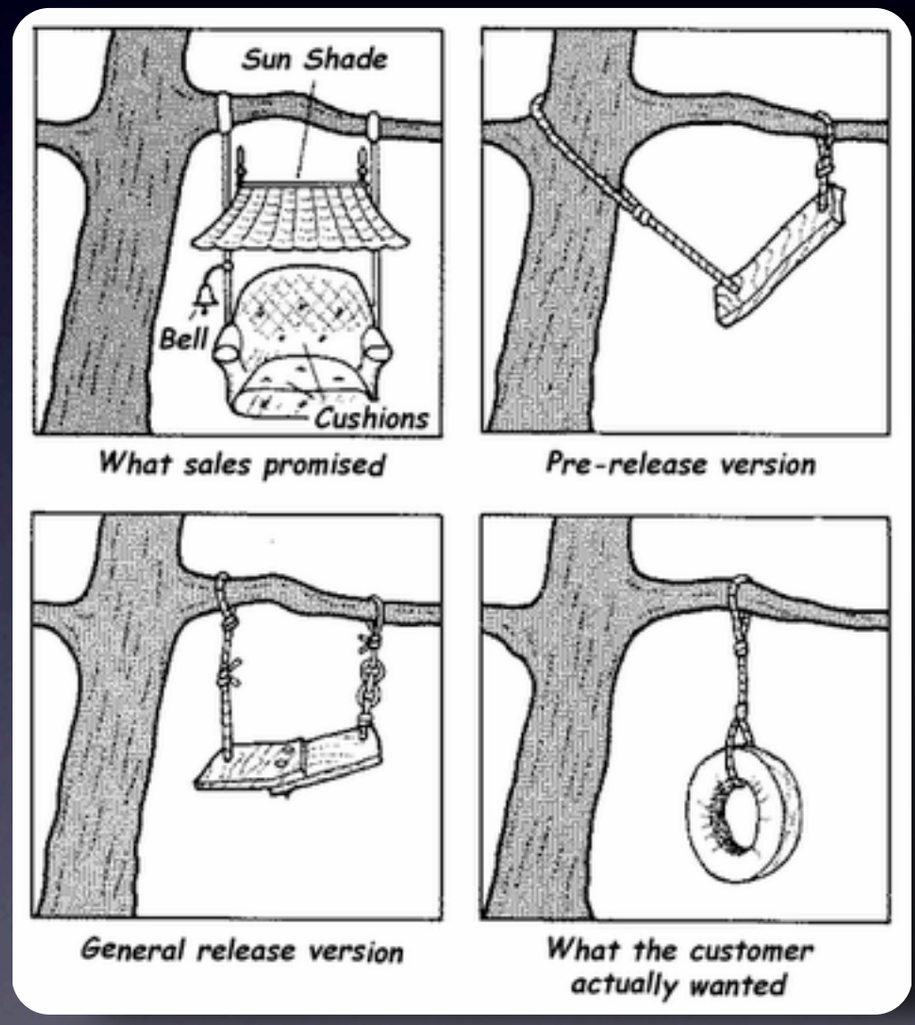
Estimates by S. Anderson, LIGO Lab

	Ligo Node (LN)	Advligo Node (AN)
Compact Binary	59,100	1,690
Continuous Wave	600	18
Burst	2000	58
Stochastic	758	22
TOTAL	62,458	1,785

- Spinning black-hole/neutron-star binaries dominate: ~760k templates
- Main Periodic analysis to be done on Einstein@Home
- 1AN = 35LN for 18 Month doubling
- 1LN = 2.2GHz of core performance
- LIGO Laboratory will supply 50%; Tier 2 centers other 50%

Learning from LIGO

- User Support:
  - About 200 (chaotic) users, many doing analysis prototyping and pushing limits of the facility - LIGO specific support
- Grid computing is still difficult for the average collaboration member to use: certificates and data management!



Enabling aLIGO Science

Applications: DQ Pipelines, Low-latency Analysis, Offline Analysis, Simulations

Software Packages: I/O Libraries, LAL Suite, MatApps, DMT, ...

Services: Auth, Support, Build/Test, Monitoring, Databases, Data Replication, h(t), ..

LIGO Data Grid: Compute & Storage Clusters, Web Services, ...

Data Center Operations: Compute & Storage Clusters, Web Services, ...

- Distributed computing: tiered data center approach with combination of US/European resources
- Technological improvements suggest we can readily meet computational needs with 1/2 from LIGO Laboratory and 1/2 from LSC Tier 2 centers
- Resource allocation and configuration for online/ offline analysis will be reviewed

Services: Support, Build/Test, Monitoring, DQ Database, Gracedb, Data Replication, h(t), ..

- Existing services span range of maturity
  - Data replication service is mature, but different ideas in LIGO and Virgo
  - DQ database worked for LIGO, but there is lots of administrative overhead
  - Deliver low-latency h(t), reduced data sets, an integrated build & test
  - Deliver robust, user friendly event database, interfaces to use distributed resources, archive service for data products

Authentication Project

- Authentication and Authorization:
  - Address need to provide easy authorization of new collaboration members
  - Deliver combined mechanism to manage collaboration tasks, e.g. which group and what activity and LDG authentication/authorization
  - Use combination of Kerberos, LDAP, Grouper, and Shiboleth, to provide single sign and ability to manage the complex set of authorizations easily

Data Replication Project

- LIGO Data Replicator:
  - moves data around the LDG before jobs are run
  - metadata catalog contains metadata information on more then 25 million files
  - each Replica Location Service (RLS) catalog between 1 and 50 million mappings
  - With eight LDR installations in the LIGO Data Grid the RLS network serves more than 300 million mappings
  - Datafind Servers: use LDR metadata to provide searchable index of data on each site using standard API

Software Packages: I/O Libraries, LAL Suite, MatApps, DMT, ...

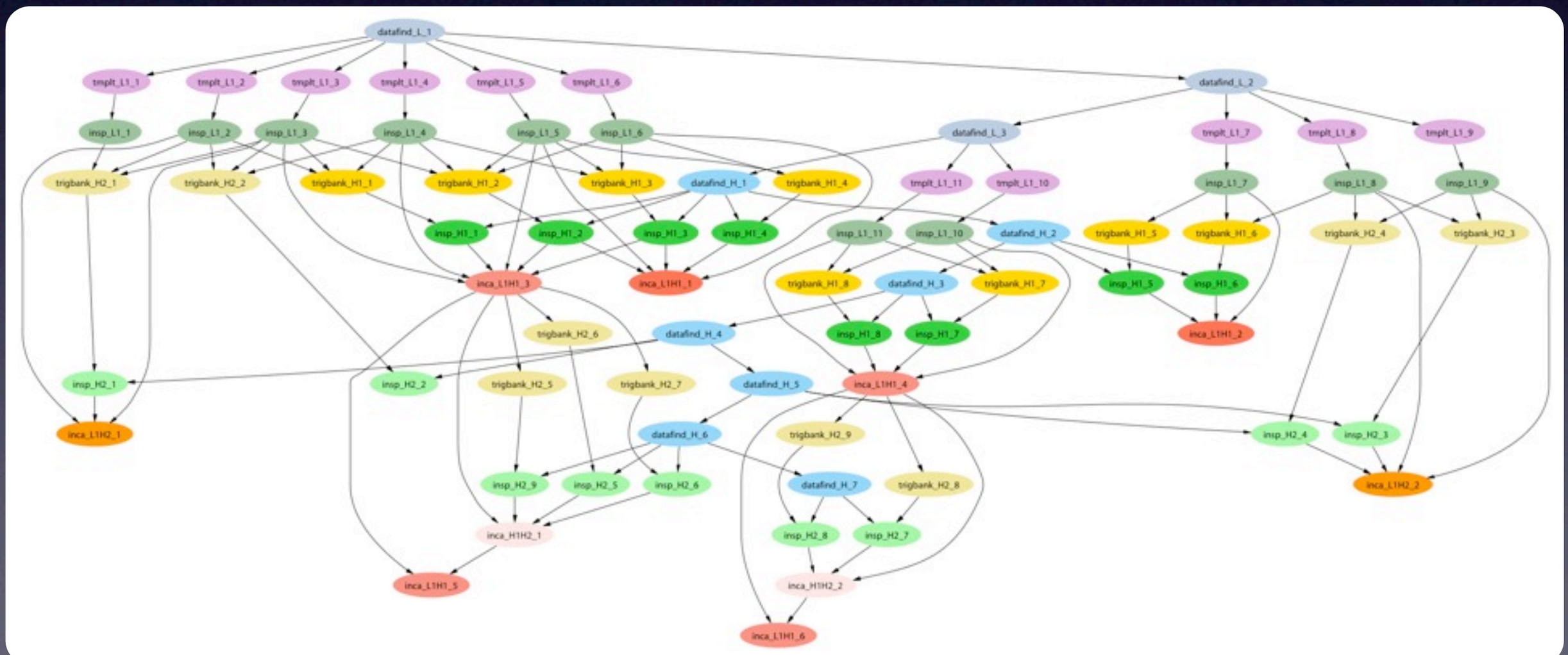
- Software stack is maturing:
  - Support different build systems in LIGO/Virgo and support different software environments enhancing code interoperability
  - Develop software release mechanisms that can be implemented for core libraries and online analysis while allowing incremental improvements
  - Provide robust APIs to new hardware, e.g. GPUs
  - Improve documentation & ease of use; provide support through help desk and training
  - Document data formats, data products and related libraries

Applications: DQ Pipelines, Low-latency Analysis, Offline Analysis, Simulations

- Regime of analysis and detector characterization groups, but ....
  - need to define data products, especially in transient searches
  - determine requirements and understand constraints of various tools
  - follow open source model to allow improvements to get back into main software stack quickly
  - document the codes for long-term viability .....

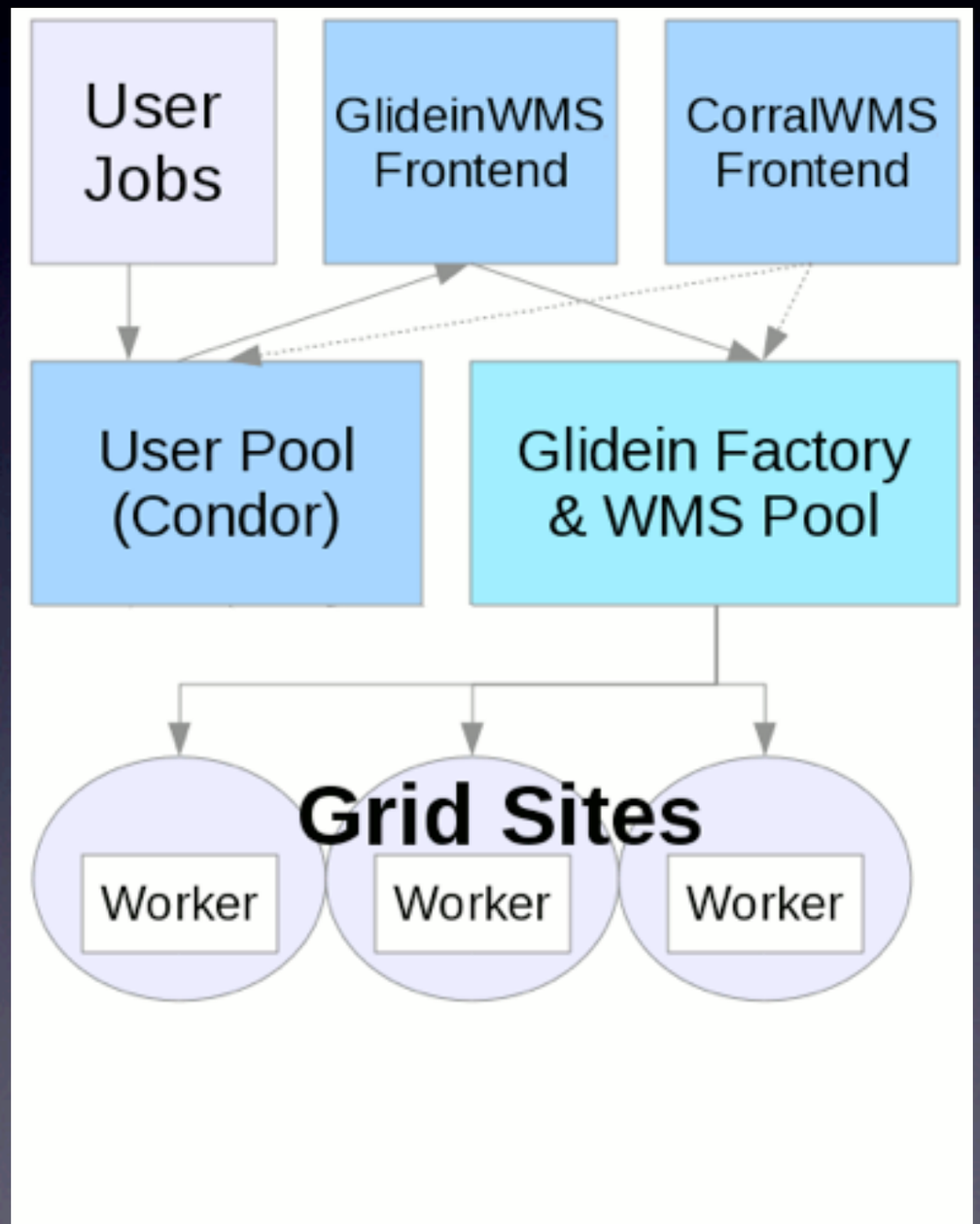
Workflow Project

- Users are used to logging into clusters which have the data on it
- Workflows run under Condor to manage dependencies
- Data management is obstacle to moving onto Grid



Workflow Project

- Condor glideins provide a nice mechanism to allow users to move from clusters to grids
- But data management is the real obstacle:
  - users are used to data access via caches at each cluster
  - but data is actually pretty small ~0.1-0.5 MB per job for long-running jobs
  - some care in workflow design to minimize overhead



Concluding remarks

- aLIGO computing planning is well underway
- keenly watching the hardware technology, but the computational and storage needs are modest
- simplify user experience in terms of authentication, grid access, data management and movement
- low-latency analysis and alert generation is a new dimension for analysis: achieved ~60 minutes in recent run, want to achieve ~10 seconds in aLIGO