



LIGO LABORATORY
MS 100-36
PASADENA CA 91125
TEL: 626.395.2129
FAX: 626.304.9834

MEMORANDUM

DATE: June 20, 2017

SUBJECT: **LIGO white paper for the NSF Large Facilities Cyberinfrastructure Workshop**

- ***A brief description of the facility, its science mission, and the community (including size, make up – number of individual, number of institutions, etc.). Please include a URL for more information.***

The Laser Interferometer Gravitational-wave Observatory (LIGO) comprises a distributed NSF facility with two 4 km x 4 km interferometers, separated by a baseline for 3,002 km, located on the DOE Hanford Nuclear Reservation north of Richland, WA and north of Livingston, LA. LIGO Laboratory is operated jointly by the California Institute of Technology and the Massachusetts Institute of Technology for the NSF under a cooperative agreement with Caltech and MIT as a sub-awardee. LIGO also includes major research facilities on the Caltech and MIT campuses.

The two gravitational wave detectors are operated in coincidence. LIGO detected gravitational waves from the inspiral and merger of a binary black hole system on 14 September 2015, heralding the opening of a new observational window on the Universe using gravitational waves to detect and study the most violent events in the cosmos.

LIGO serves the worldwide gravitational wave community through the LIGO Scientific Collaboration, consisting of over 40 institutions in 15 countries. This international collaboration comprises about 1,100 members. LIGO also has MOUs covering joint operations with the EU Virgo Collaboration and the Japanese KAGRA Collaboration.

More information about LIGO may be obtained at the following URL:
<https://www.ligo.caltech.edu>

- ***A description of the key products/services of the facility (data, software, services, etc.)?***

The key data product generated by LIGO is a time series recording relative changes in length between the two 4km arms of each LIGO interferometer. These strain measurements (~3 TByte/y) record audio-frequency perturbations in the local space-time metric at each Observatory at the level of 1 part in 10^{22} . This is the primary observable from the LIGO experiment, recording the signature of gravitational waves



passing through each detector. To inform data analysis efforts searching the strain data for gravitational waves, and to understand and improve the performance of the LIGO instruments, an additional ~200k channels of environmental monitors and internal instrument channels are recorded (1.5 PByte/y). The strain data are distributed in low-latency (seconds) to computing clusters running analysis pipelines to generate gravitational-wave triggers for external Astronomical observations for transient events on a timescale of 1 minute. The bulk data are locally archived at each LIGO Observatory and distributed over the Internet to a central data archive on a timescale of 30 minutes. The central data archive currently holds 7 PByte of LIGO observations in perpetuity.

LIGO data analysis software is released using native Linux packaging (.rpm and .deb) and pre-installed on dedicated computing resources via standard Linux software repositories. For computing on shared resources the software is distributed via the CERN Virtual Machine Filesystem (CVMFS) and containerized with Docker, Singularity, or Shifter. Similarly, the key science data are pre-staged on dedicated computing resources ahead of analysis, and distributed to shared computing resources via CVMFS or GridFTP as needed by computing tasks. Metadata that describe LIGO observations and candidate signals from data analysis are stored in databases with custom tools for ingestion and querying.

- ***A brief description (including a figure) of the facility CI (e.g., its architecture, key services/components, underlying infrastructure), how it is deployed/distributed, and its operation. What is the median age since deployment of the key CI components.***

LIGO data analysis computing overwhelmingly consists of embarrassingly parallel workflows executed on high-throughput (HTC) resources. The majority of LIGO computing is provided by internal LIGO Scientific Collaboration (LSC)-managed clusters, but a growing fraction is provided by external shared resources. These resources are integrated into LIGO's computing environment via the Open Science Grid, and consist of a variety of dedicated and opportunistic campus, regional, and national clusters, Virgo scientific collaboration resources, and XSEDE allocations.

LIGO relies on HTCondor for its internal job scheduling, and uses both DAGMan and the Pegasus WMS for large-scale workflow management on top of HTCondor. In addition, LIGO uses the BOINC infrastructure to manage its single largest data analysis task (the search for continuous wave signals) via Einstein@Home running on volunteer computers as a screen saver. For Single Sign-On and other Identity



and Access Management functions, LIGO relies on Shibboleth, Grouper, InCommon, and CILogon. The underlying authentication infrastructure is built on Kerberos and authorization information if reflected in LDAP.

For distributed data management, LIGO relies on CVMFS, StashCache/Xrootd, Globus GridFTP, and a variety of in-house CI tools and services to complement and integrate these tools.

