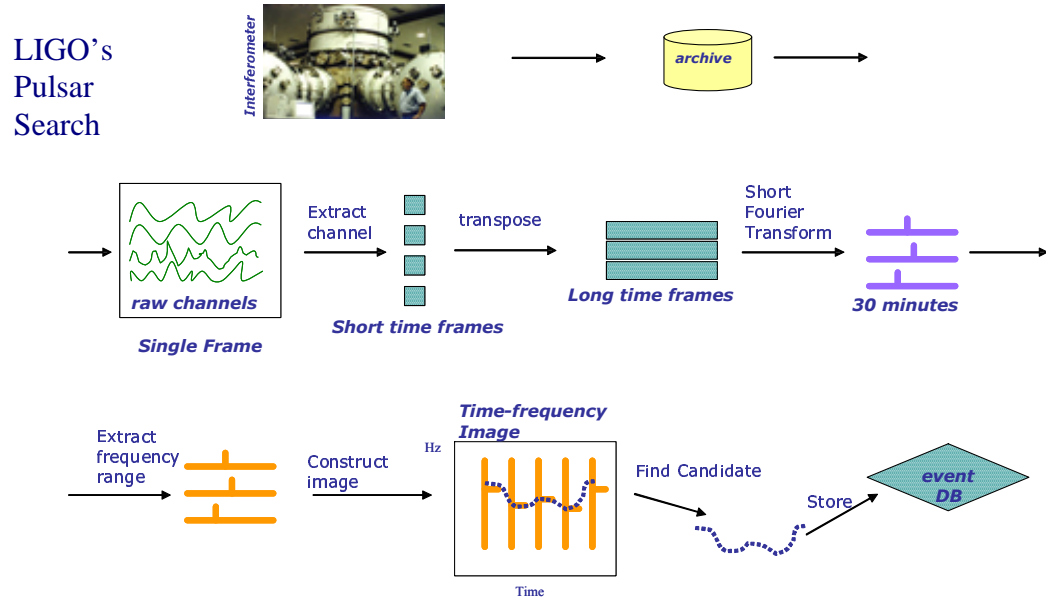


The LIGO Story

The Laser Interferometer Gravitational Wave Observatory (LIGO) is making great strides towards performing scientifically significant data analysis using Grid resources. As part of the collaboration between the LIGO and the GriPhyN projects, we have focused on a specific LIGO problem: the pulsar search (shown below). The data needed to conduct the search is a long stretch (~4 months, 2×10^{11} points) of a single channel—the gravitational wave strain channel. The power spectra of the small segments are stacked to make a large frequency-time image, perhaps 4×10^5 on each side. The pulsar search consists of searching for coherent signals in this image. A source would appear on the frequency-time image as a wavering line, whose frequency might be 1 kHz, but modulated by a few parts in 10^6 over periods of 1 day and a few parts in 10^4 over periods of 1 year. In addition, if the source exhibits any secular variations due to slowing down of its rotational period, these will be encoded in the data, as well.



The pulsar search is very compute and data intensive and requires more resources than those available in the LIGO Scientific Collaboration. In order to take advantage of the Grid resources, we had to integrate LIGO's existing data analysis into the Grid environment. The LIGO Data Analysis System (LDAS) can perform a wide range of sophisticated and computationally intensive data analysis. We have developed an infrastructure in which LDAS can be accessed as a Grid resource, and in which LDAS can schedule its jobs on the Grid. To achieve integration of LDAS into the Grid environment, we have designed and implemented Globus interfaces both for data staging (via GridFTP) and computation scheduling (via GRAM). These interfaces also make use of the Globus Security Infrastructure to provide secure access to resources.

The pulsar search conducted at SC 2002 uses LIGO's data collected during the first scientific run of the instrument and will target a set of 1000 locations of known pulsar as well as random locations in the sky. The results of the analysis will be published via LDAS to the LIGO Scientific Collaboration and thus will become accessible to LIGO scientists.

LIGO

Participants:

Stuart Anderson
Kent Blackburn
Masha Barnes
Philip Charlton
Phil Ehrens
Ed Maros
Greg Mendell
Albert Lazzarini
Mary Lei
Isaac Salzman
Peter Shawhan

California Institute
of Technology

Scott Koranda
University of
Wisconsin
Milwaukee and
NCSA

Maria Alessandra
Papa
Albert Einstein
Institute, Germany

Alicia Sintes
Balearic Islands
University, Spain



The Pegasus Story



Pegasus Participants:

Amit Agarwal
Shishir Bharathi
James Blythe
Ewa Deelman
Yolanda Gil
Carl Kesselman
Gaurang Mehta
Sonal Patil
Srividya Rao
Gurmeet Singh
Marcus Thieboux
Karan Vahi
Information Sciences Institute USC

For more information please visit:

GriPhyN:
www.griphyn.org

Pegasus:
www.isi.edu/~deelman/pegasus

MCS:
www.isi.edu/~deelman/MCS

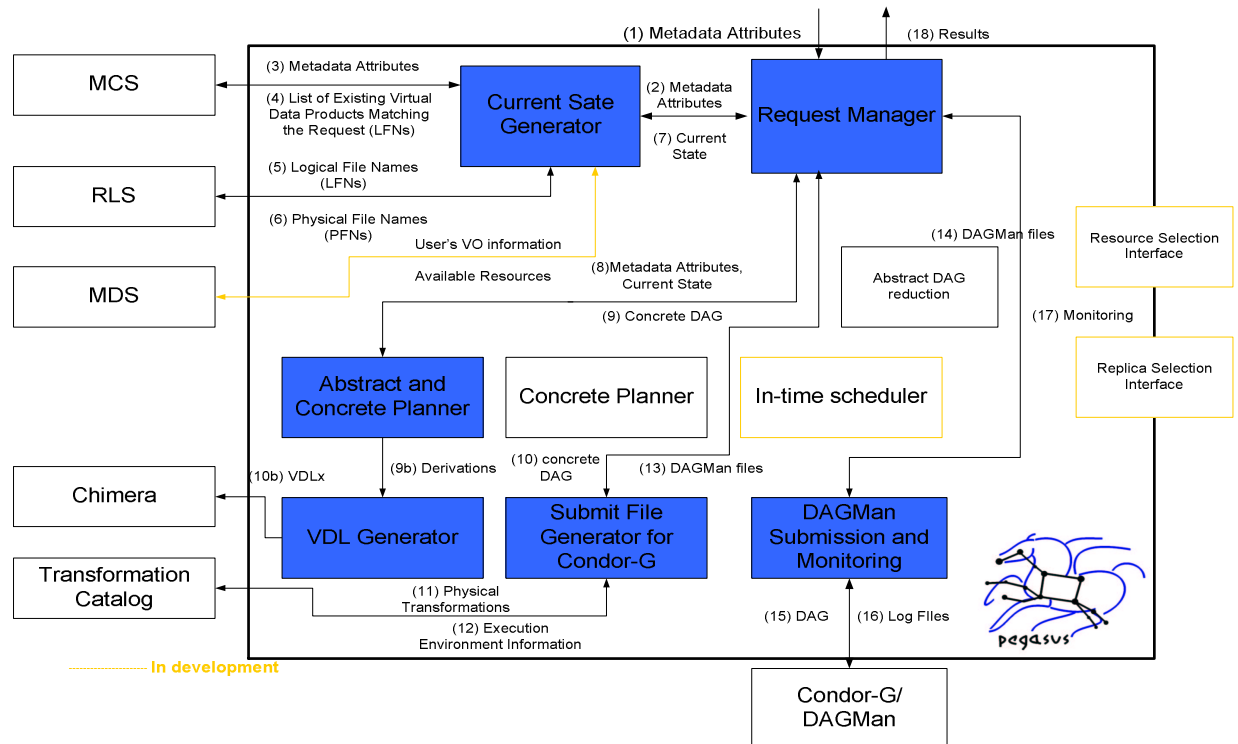
RLS:
www.isi.edu/~annc/RLS

Chimera:
www-unix.griphyn.org/chimera

DAGMan:
www.cs.wisc.edu/condor/dagman/

Pegasus, which stands for Planning for Execution in Grids, was developed at ISI as part of the GriPhyN project. Pegasus is a configurable system that can map and execute complex workflows on the Grid. Pegasus has been integrated with the GriPhyN Chimera system. In that configuration, Pegasus receives an abstract workflow (AW) description from Chimera, produces a concrete workflow (CW), and submits it to DAGMan for execution. AW describes the transformations and data in terms of their logical names. CW, which specifies the location of the data and the execution platforms, is optimized by Pegasus from the point of view of Virtual Data. If data products described within AW are found to be already materialized (via queries to the Globus Replica Location Service (RLS)), Pegasus reuses them and thus reduces the complexity of CW.

Pegasus can also be configured to perform the generation of the abstract workflow based on application-level metadata attributes. Given attributes such as time interval, frequency of interest, location in the sky, etc., Pegasus is currently able to produce any virtual data products present in the LIGO pulsar search. The figure below shows Pegasus configured for such a search.



Pegasus uses the Metadata Catalog Service (MCS), newly developed at ISI, to perform the mapping between application-specific attributes and logical file names of existing data products. AI-based planning technologies are used to construct both the abstract and concrete workflows. Pegasus also contains a Virtual Data Language generator that can populate the Chimera catalog with newly constructed derivations. Finally, Pegasus generates the necessary submit files and sends the CW to DAGMan for execution. The LIGO pulsar search is performed using the Pegasus system, LDAS and compute and storage resources at Caltech, University of Southern California, University of Wisconsin Milwaukee, University of Florida, and NCSA.