LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Document Type	LIGO-T070219-00-Z	2008/09/04
	LIGO Data Grid Reference Operating System	
	Stuart Anderson, Warren Anderson, Kent Blackburn, Patrick Brady, Duncan Brown, Steve Fairhurst, Sam Finn, Erik Katsouvanidis, Greg Mendel, Tom Nash, Reinhard Prix	

Distribution of this draft:

LIGO Scientific Collaboration

California Institute of Technology
LIGO Project - MS 51-33
Pasadena CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project - MS NW22-295
Cambridge, MA 01239
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

WWW: http://www.ligo.caltech.edu/

Contents

1 Summary

The LIGO Data Grid (LDG) is a federation of computer clusters run by the LIGO Scientific Collaboration (LSC) to enable astrophysical data analysis of LIGO and GEO science data. To help facilitate this scientific goal it is desirable to provide a standard computing environment for the development and maintenance of software tools as well as the ease of moving analysis tasks from one physical cluster to another within the LDG. To this end the LSC Computer Committee (CompComm) has selected CentOS5 as the next reference platform to replace the current reference of FC4 once the current S5 science run is completed. In addition, Debian Etch has been selected as a secondary platform. In case of limited resources available, CentOS5 support will come first, but Debian will be supported as well, possibly requiring a somewhat longer timescale.

The definition of the LDG Reference Platform is the common OS that all LDG software should support at a minimum. This provides at least one OS that can support the full set of LSC software in a single environment. It is a recommendation to the individual LDG Cluster owners to use either the reference platform (preferred) or the secondary platform whenever possible, but not a strict requirement. The Reference Platform is also the OS that the LSC will request the highest level of support for from external entities which provide important software and hardware for the LSC. Examples of external software include the primary LDG job management system, currently Condor, and the Grid Middleware used by the LSC, currently Globus as distributed through the Virtual Data Toolkit (VDT).

2 Introduction

The reference operating system (OS) platform for the LDG during the LIGO S5 run (November 2004 thru October 2007) was Fedora Core 4 (FC4). Given that FC4 is no longer actively maintained or supported by an external entity the CompComm established the following working group in April 2007 to evaluate the technical merits and costs of various OS's to adopt as the next LDG reference platform once S5 is completed:

- Paul Armor: UW-Milwaukee Tier 2 Cluster

- Lisa Bogue: CDS Interface

- Duncan Brown (Chair): Analysis Software

- Erik Espinoza: LIGO Tier 1 Cluster

- Kevin Flasch: LDR/Globus Middleware

- Steffen Grunewald: GEO Tier 1 and 2 Clusters

- Ben Johnson: Observatory Clusters

- Tyler Petire: Penn State Tier 2 Cluster

- Henning Fehrmann: GEO Tier 1 Cluster

- Carsten Aulbert: GEO Tier 1 Cluster

- Gerald Davies: GEO Tier 1 Cluster, Virtualization Support.

This working group evaluated the following operating systems:

1. Commercial GNU/Linux distributions:

    (a) Red Hat Enterprise Linux
    (b) SuSE Enterprise Linux

2. Free (community-supported) GNU/Linux distributions:

    (a) CentOS
    (b) Scientific Linux
    (c) Debian
    (d) Ubuntu
    (e) Fedora Core

3. Non-Linux operating Systems:

    (a) Solaris
    (b) Free BSD

Their report narrowed down the selection to two choices: CentOS5 and Debian/Ubuntu. The full report is LIGO document T070102.

3 Discussion

The primary finding from the technical working group evaluating the next OS for the LDG was that there are 2 viable choices that could serve the prime functions of the LDG: CentOS and Debian. In short, the main finding in favor of Debian is a richer set of natively supported packages and in favor of CentOS is better support for Condor and QFS support from Sun Microsystems.

In order to further down select from these 2 choices recommended by the technical working group two additional investigations where completed: a technical discussion with the Condor development team on the current and planned support model for Condor on different Linux distributions, and an evaluation of current Linux support from major hardware vendors of interest.

The findings from the Condor team are that Condor development and support is strongly oriented towards RedHat Enterprise Linux (RHEL) and its derivates. This is motivated by: the large High Energy Physics commitment to both Condor and Scientific Linux (RHEL derivative), a large fraction of Condor paying customers running RHEL, a new strategic relationship established this year between RedHat and Condor at the University of Wisconsin in Madison, and the host department and institution for the Condor team run large Condor pools on Scientific Linux and CentOS. However, it is also apparent that the technical challenges of getting Condor to run fully on another

Linux platform, such as Debian Etch, are primarily limited to matching the combination of kernel, libc, and gcc. It is also possible to compile fully functional Condor applications (including remote I/O and checkpointing) on a Condor supported platform and run the statically linked executables on a wider variety of Linux distributions. Therefore, supporting Condor on any LDG clusters running Debian Etch may not be too difficult once Condor has been ported and is fully supported on one of the newer gcc version 4 platforms, such as the planed RHEL support.

The investigation of Hardware support was carried out by Bruce Allen who found the following bias towards RHEL:

1. IBM: supports Novell Linux and Redhat on all servers

2. Sun: supports Redhat and SuSE. Also works with Canonical to certify and support x64 and CoolThreads on Ubuntu

3. HP: supports Redhat on all systems and SuSE/Novell on most systems. They also additionally support Debian, and Oracle EL on some systems. They are also offering some Red Flag and Mandriva support.

4. Dell: supports Redhat and Novell/SuSE. Coming soon will offer Ubunto on select desktop and notebook systems.

5. Fujitsu: supports Redhat and Novell/SuSE

6. Intel: mixed support for Redhat, Red Flag, SuSE/Novell, Mandriva, CS2C. Debian Sarge

4   Conclusion

Based on the findings of the technical evaluation working group which identified CentOS and Debian as viable alternatives the CompComm has selected CentOS 5 as the next reference platform with Debian Etch as the secondary platform that will be supported as resources permit. This decision was also based on discussions with the Condor team, previous experience with providing missing packages from the core OS via the LSCSOFT repository, support for QFS, and the apparent stronger commitment for RHEL from the hardware manufactures. It is also important to note that if the LDG ends up running both CentOS and Debian, that will be a practical improvement over the S5 time period where the LDG was comprised of 3 different operating systems. Furthermore, supporting a secondary platform (as resources permit) allows a fall-back plan for our computing infrastructure in case the future development of one distribution does not turn out as hoped.