



# Los Alamos National Laboratory

## *MySQL Scales with 7 Terrabytes of Data*

Los Alamos National Laboratory is one of 28 Department of Energy (DOE) laboratories across the country, and is the largest institution and largest employer in Northern New Mexico. The laboratory's annual budget is approximately \$1.2 billion. The Los Alamos Research Library focuses on providing products and services to increase the efficiency of scientists and research staff in exploiting their information repositories.

The Research Library had 8 databases, containing scientific articles as well as meta data for those articles such as abstracts, biographies, and references. Each of the databases was from different providers and data was stored in unique formats making it virtually impossible for the research staff to effectively search for content across all repositories.

Los Alamos needed a system that would enable scientists and researchers to search for articles and authors more efficiently, without having to know the details of each database.

To solve this problem, Los Alamos Research Library developed *SearchPlus*, a comprehensive distributed database containing more than 55 million scientific journal articles that can be accessed and searched from a browser, any time, anywhere. Scientists can now access and easily



*"MySQL has proven itself to be the ideal solution for SearchPlus by combining high-performance, superior reliability, security and replication."*

**Mariella Di Giacomo**  
Los Alamos National Laboratory  
Research Library



## MySQL Delivers Robust, Scalable and Secure System

search a consolidated set of data using a simple interface to do research, publish articles, and compete for funds in highly competitive fields. MySQL is the high-performance database that enables Los Alamos Labs to deliver this robust, scalable and secure *SearchPlus* system.

### Requirements

#### Security

Los Alamos National Laboratory Research Library not only needed to provide search services in a secure environment behind a firewall, but also need to make these services available through a web application outside the firewall for external customers. In addition, user authentication and authorization was required.

#### Performance

One of the primary goals was delivering responsive and interactive search services. The Research Library data centers have particularly demanding size and speed requirements. The research library needed a database that could store article and citation meta data in common format for over 500 million entries.

#### Reliability

The Research Library needed a reliable system that was fault tolerant and would not result in any loss of data. Not only do internal scientists and research staff rely on the Research Library but several institutions including Stanford, DOE Labs, and the Air Force Research Library Consortium pay for access to *SearchPlus* services, so reliability is critical.

#### Scalability

The Research Library needs to continually add new data. Building a scalable and adaptable system capable of handling weekly updates, new data sources, formats, and content was critical to being able to offer superior services.



## The *SearchPlus* System

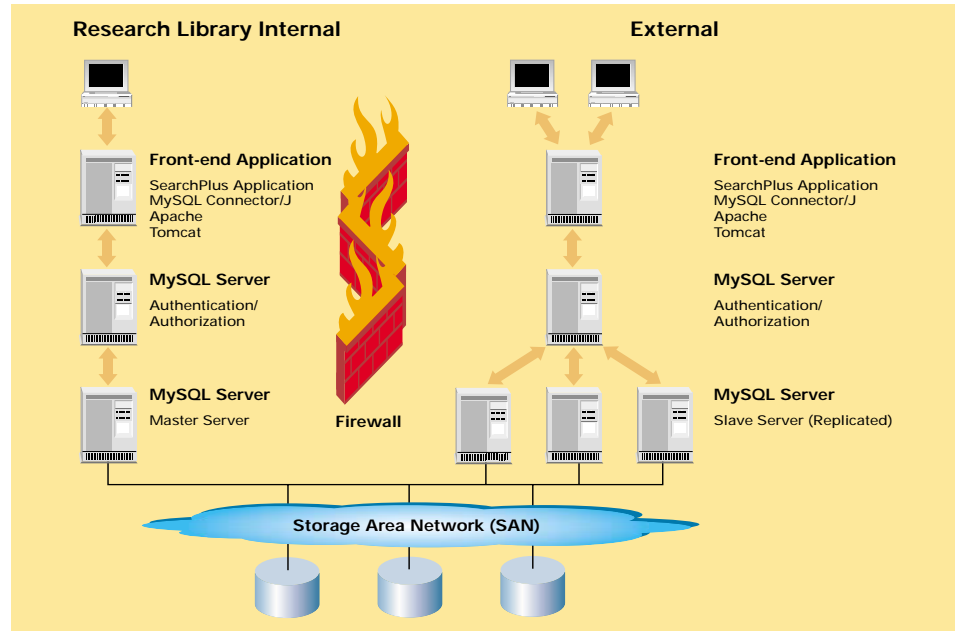
*SearchPlus* provides basic and advanced search capabilities with an easy to use browser interface for cross-searching databases, enabling researchers to:

- ◆ Search and View Article Abstracts
- ◆ Search and browse citation information
- ◆ View how many times authors are cited
- ◆ Browse cited authors and search cited references
- ◆ Access the full-text article

Additional *SearchPlus* features include:

- ◆ Login to set personal settings
- ◆ Mark records to print, e-mail or download records
- ◆ Set up alerts to stay current with the literature
- ◆ View your search history
- ◆ Easily do multiple searches for a given period

MySQL stores all the bibliographic data for scientific publication in a common XML format for indexing and retrieval.



*Using MySQL, Los Alamos National Labs accelerates research with fast access to more than 55 million scientific journal articles.*

The Research Library is running multiple MySQL servers both inside and outside the firewall that serve as authentication/authorization databases, author databases, and citation (cited references) databases. A single unified interface allows users to login and search the databases to dynamically build bibliography and citation counts enabling researchers to see how many times authors are cited. Cited references are linked to the scientific articles for easy lookup.

The Research Library relies on MySQL for the following primary reasons:

- ◆ **High Performance** – MySQL stores over 1.4 billion rows of data. The MyISAM storage engine is the technology that enables *SearchPlus* to be extremely responsive.
- ◆ **Security** – *SearchPlus* uses MySQL to provide secure login by authenticating and authorizing users both internally and externally.
- ◆ **Replication** - the Research Library replicates data outside the firewall in a master-slave configuration so users both inside and outside the firewall can search and browse the database.

## Technical Environment

### SearchPlus Architecture

*Nodes:* 12 Processing Nodes  
46 Processors

*RAM:* 234GB of Main Memory  
~ 7TB of Disk Storage

### MySQL Servers

*Nodes:* 8 Processing Nodes

*Hardware:* One Sun V480 as the Master Server and another as the Slave Server.

*OS:* Linux

*RAM:* 65 GB RAM

*Language:* Java

*Web Server:* Apache/Tomcat

*Database:* MySQL Server 4.0

*Storage:* RAID and SAN

*Database Size:* Each server hosts a MySQL DB of approximately

- 500GB
- 1.4 billion rows

## About MySQL AB

MySQL AB develops and markets a family of high performance, affordable database servers and tools. The company's flagship product is MySQL, the world's most popular open source database, with more than 5 million active installations. Many of the world's largest organizations, including Yahoo!, Sabre Holdings, Cox Communications, The Associated Press and NASA, are realizing significant cost savings by using MySQL to power Web sites, business-critical enterprise applications and packaged software. MySQL AB is a second generation open source company, with dual licensing that supports open source values and methodology in a profitable, sustainable business.

For more information about MySQL, please go to [www.mysql.com](http://www.mysql.com).



The World's Most Popular Open Source Database

## MySQL Worldwide Offices

### North America Headquarters

2510 Fairview Avenue East  
Seattle, WA 98102 USA  
+1-425-743-5635 P  
+1-425-671-0771 F

### France

+33-(0)1-43-077-099

### Germany, Austria, Switzerland

+49-(0)7022-9256-30

### Worldwide Headquarters

Bangårdsgatan 8  
S-753 20 Uppsala  
Sweden  
+46-730-234-111 Sales

### Spain, Portugal, Latin America

+1-425-373-3434

### Finland

+358-(0)-9-2517-5553