

Grid Developers: Organizations and Projects Worth Watching

Level: Introductory

Edna Nerona (edna@legacystudios.biz), Consultant, Legacy Studios

14 June 2007

Learn more about the innovative projects and leading organizations that are advancing Grid development today.

Introduction

As you previously read, grid computing pools and maximizes the value of computing resources. You also got an understanding of the open source tools and the variety of resources available to you in the previous two articles. Now it's time to familiarize you with the production projects and organizations that are shaping the future of grid computing.

This article provides a comprehensive list of current projects in diverse areas such as cancer research, astronomy, physics, just to name a few. We will also cover toolkits, security, and data management. These were taken from a variety of online sources to introduce programmers, administrators and new users to specific information and projects related to using, deploying and developing on grid infrastructure.

This list is updated periodically. Please help us improve it by providing your comments below.

Production Grid Organizations

This track will introduce you to production grid organizations, the problems they are solving, and how their influencing grid technologies. Production Grid deployments fall into various categories of grids – general-purpose grids, scientific and community grids, nation-wide grids, regional and university grids.

National and International General Purpose Grids

DEISA

<https://www.deisa.org/>

As a consortium of leading national supercomputing centers, DEISA currently deploys and operates a secure, production quality, distributed supercomputing environment. By enhancing and reinforcing European capabilities in the area of high performance computing, the research infrastructure facilitates scientific discoveries across a wide range of science and technology fields. DEISA uses a deep integration of existing national high-end platforms, with a dedicated network and support by innovative system and grid software.

DutchGrid

<http://www.dutchgrid.nl/>

Established in 2000, the Dutch Grid has a many successful integrated efforts and initiatives that span a wide range of scientific collaborations. As an open platform for academic and research grid computing, DutchGrid provides globally recognized identity certificates to grid users in the Netherlands. The DutchGrid CA is fully project-neutral. Any not-for-profit researcher and academic user can obtain personal and server or host certificates for use with grid applications.

Enabling Grids for E-science (EGEE)

<http://www.eu-egee.org>

The Enabling Grids for E-science (EGEE) project brings together scientists and engineers from more than 90 institutions in 32 countries world-wide to provide a seamless Grid infrastructure for e-Science that is available to scientists 24 hours-a-day. The EGEE Grid consists of over 30,000 CPU available to users 24 hours a day, 7 days a week, in addition to about 5 Petabytes (5 million Gigabytes) of storage, and maintains 30,000 concurrent jobs on average. Having such resources available changes the way scientific research takes place. The end use depends on the users' needs: large storage capacity, the bandwidth that the infrastructure provides, or the sheer computing power available. EGEE is a four-year project funded by the European Commission.

Grid5000

<https://www.grid5000.fr>

The purpose of Grid'5000 project is a highly reconfigurable, controllable and monitorable experimental grid platform that grid researchers can use as a testbed for experiments in all the software layers between the network protocols and up to applications. Grid5000 brings together 9 sites geographically distributed in France featuring a total of 5000 CPUs. These areas include Bordeaux, Grenoble, Lille, Lyon, Nancy, Orsay, Rennes, Sophia-Antipolis, and Toulouse.

LA Grid

<http://latinamericangrid.org/>

Pronounced "lah grid," the LA Grid is the first-ever comprehensive computing grid to connect faculty, students, and researchers from institutions across the United States, Latin America and Spain to collaborate on complex industry applications for business and societal needs in the context of health. In addition to universities, LA Grid has partnered with industries across the globe, enhancing the innovations in many areas including healthcare, life sciences and hurricane disaster, life sciences and disaster mitigation.

Open Science Grid

<http://www.opensciencegrid.org/>

The Open Science Grid is a distributed computing infrastructure for scientific research. The OSG Consortium's unique alliance of universities, national laboratories, scientific collaborations and software developers brings petascale computing and storage resources into a uniform shared cyberinfrastructure.

TeraGrid

<http://teragrid.org/>

TeraGrid is an open scientific discovery infrastructure funded by the National Science Foundation. Combining leadership class resources at nine partner sites, TeraGrid creates an integrated, persistent computational resource. Interconnected via a high-speed gigabits/second dedicated national network, TeraGrid provides more than 150 teraflops of computing power and nearly 2 petabytes of rotating storage, numerous scientific data collections, specialized data



analysis tools, scientific gateways and user portals to simplify access to valuable resources, and visualization resources.

Scientific & Community Grids

AstroGrid

<http://www2.astrogrid.org/>

AstroGrid is an open source project built to create a working Virtual Observatory (VO) for UK and International astronomers. Funded by the UK government, AstroGrid works closely with other VO projects worldwide through the International Virtual Observatory Alliance (IVOA). As a leading member of this community, AstroGrid provides internationally recognized interface standards that are emerging to promote scientific integration of astronomical data and processing resources worldwide.

caBig Grid

<http://cabig.nci.nih.gov>

The cancer Biomedical Informatics Grid, or caBIG, is a voluntary network or grid connecting individuals and institutions to enable the sharing of data and tools, creating a World Wide Web of cancer research. The goal is to speed the delivery of innovative approaches for the prevention and treatment of cancer. The infrastructure and tools created by caBIG also have broad utility outside the cancer community. caBIG is being developed under the leadership of the National Cancer Institute's Center for Bioinformatics.

International Virtual Data Grid Laboratory

<http://www.ivdgl.org>

The International Virtual Data Grid Laboratory (iVDGL) is a global Data Grid that will serve forefront experiments in physics and astronomy. Its computing, storage and networking resources in the U.S., Europe, Asia and South America provide a unique laboratory that will test and validate Grid technologies at international and global scales. Sites in Europe and the U.S. will be linked by a multi-gigabit per second transatlantic link funded by the European DataTAG project.

World Community Grid

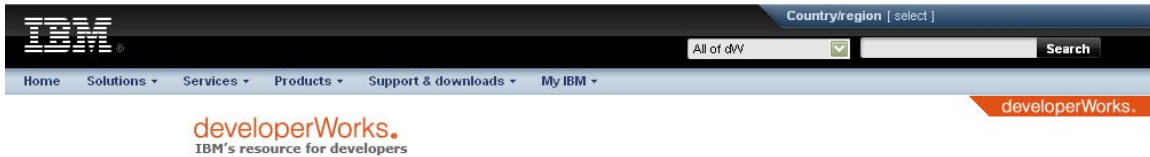
<http://www.worldcommunitygrid.org>

World Community Grid's mission is to create the world's largest public computing grid to tackle projects that benefit humanity. The success of the World Community Grid depends upon individuals collectively contributing their unused computer time to change the world for the better. World Community Grid is making technology available only to public and not-for-profit organizations to use in humanitarian research that might otherwise not be completed due to the high cost of the computer infrastructure required in the absence of a public grid.

WorldWide LHC Computing Grid

<http://lcg.web.cern.ch/LCG/>

The Large Hadron Collider Computing Grid is designed to handle the unprecedented quantities of data that will be produced by experiments at CERN's Large Hadron Collider (LHC) from 2007 onwards. The computational requirements of the experiments that will operate at the LHC are enormous. Some 12-14 petabytes of data will be generated each year, the equivalent of more than 20 million CDs. Analyzing this data will require the equivalent of 70,000 of today's fastest PC computers. The LCG will meet these needs by deploying a worldwide computational Grid,



integrating the resources of scientific computing centers spread across Europe, America and Asia into a global virtual computing service.

US Regional Grids

Northwest Indiana Computational Grid

<http://www.nwicgrid.org/>

Northwest Indiana Computational Grid (NWICG) is a partnership of researchers and educators from Purdue University-West Lafayette, Purdue University-Calumet, and the University of Notre Dame. With a focus on national science and research initiatives, NWICG creates cyberinfrastructure that supports the solution of breakthrough level problems, and enabling continuing world-class advances in the underlying technologies of high performance computing. They are developing a scalable, high speed, high bandwidth, and science driven computational grid for Northwest Indiana across the three universities in collaboration with the Department of Energy's Argonne National Laboratories.

SURAGrid

http://www.sura.org/programs/sura_grid.html

Southeastern Universities Research Association (SURA) is a consortium of organizations collaborating and combining resources to help bring grid technology to the level of seamless, shared infrastructure. The SURAGrid focuses on direct access to a rich set of distributed capabilities for participating research and education communities. SURAGrid promotes the development of contributed resources, project-specific tools and environments, highly specialized or HPC access, and gateways to national and international cyberinfrastructure.

TIGRE

<http://www.hipcat.net/Projects/tigre>

The mission of the Texas Internet Grid for Research and Education (TIGRE) project is to create a computational grid that brings together computing systems, storage systems and databases, visualization laboratories and displays, and even instruments and sensors across Texas. By enhancing the computational capabilities for Texas researchers in academia, government, and industry by integrating massive computing power, TIGRE's hopes to aid in the advancement of biomedicine, energy and the environment, aerospace, materials science, agriculture, and information technology.

Open-Source Grid Projects

These Grid projects cover a diverse set of areas ranging from grid infrastructure toolkits, middleware toolkits, data tools, security and more. The following represent some fast moving projection grid projects and tools. Visit these links often to keep up to date on how they are leading the progress in grid technology.

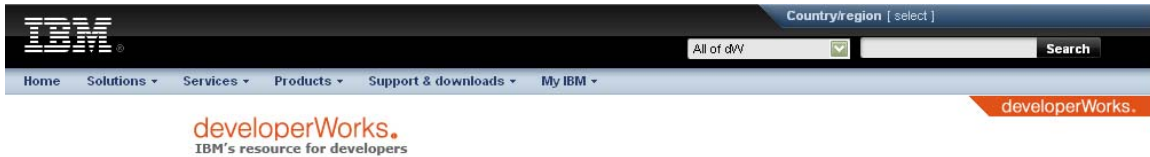
Grid Infrastructure Projects

Open source grid infrastructure projects that can help you set up your own grid.

BOINC

<http://boinc.berkeley.edu/>

Berkeley Open Infrastructure for Network Computing (BOINC) is a software platform for projects, like distributed.net and SETI@home, that use millions of volunteer computers as a



parallel supercomputer. Source code is available for the platform, and interested C++ developers are encouraged to help develop the platform code. BOINC is currently supported on Windows, Linux, UNIX, and Mac OSX platforms. CPU platform requirements may vary among project clients using BOINC.

UNICORE

<http://www.unicore.eu/>

UNICORE (Uniform Interface to Computing Resources) offers a ready-to-run Grid system including client and server software. UNICORE makes distributed computing and data resources available in a seamless and secure way in intranets and the Internet. The UNICORE design focuses several core principles: seamless access to heterogeneous environments; security; site autonomy; a powerful GUI clients that provides ease of use; and quick start bundles that allow for simple installation.

Grid Middleware Projects

The following projects have successfully provided US and international projects with the advanced tools to easily access numerous grid functionalities such as computation, visualization and storage resources. You can interact with various grids or have one customized to work with your own grid.

gLite

<http://glite.web.cern.ch/glite/>

gLite is the next generation middleware for grid computing, born from the collaborative efforts of more than 80 people in 12 different academic and industrial research centers as part of the EGEE Project. gLite provides a bleeding-edge, best-of-breed framework for building grid applications tapping into the power of distributed computing and storage resources across the Internet.

NAREGI

http://www.naregi.org/index_e.html

NAREGI, the National Research Grid Initiative in Japan, focuses on the research and development of grid middleware so that a large-scale computing environment can be implemented for widely-distributed, advanced research and education.

Ninf-G

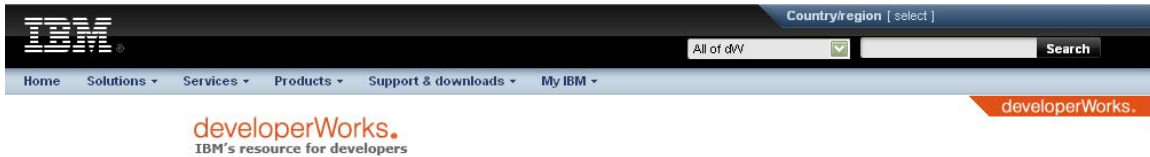
<http://ninf.apgrid.org/>

Ninf is a Japanese project developing programming middleware which enables users to access various resources such as hardware, software and scientific data on the Grid with an easy-to-use interface. Ninf-G is an open source software which supports development and execution of Grid-enabled applications using Grid Remote Procedure Call (GridRPC) on distributed computing resources.

NorduGrid Middleware

<http://www.nordugrid.org/middleware/>

The NorduGrid middleware (or Advanced Resource Connector, ARC) is an open source software solution distributed under the GPL license, enabling production quality computational and data Grids. ARC provides a reliable implementation of the fundamental Grid services, such as information services, resource discovery and monitoring, job submission and management, brokering and data management and resource management. Most of these services are provided



through the security layer of the GSI. The middleware builds upon standard Open Source solutions like the OpenLDAP, OpenSSL, SASL and Globus Toolkit® (GT) libraries.

OGSA-DAI

<http://www.ogsadai.org.uk/>

The OGSA-DAI project focuses on the development of middleware to assist with the access and integration of data from separate sources through the grid. OGSA-DAI is funded as one of three major UK Open Middleware Infrastructure Institute UK that provide software and support to the UK e-Science community and its international partners. The project works closely with the Globus, OMII-Europe, NextGRID, SIMDAT and BEinGRID groups, ensuring the OGSA-DAI software works in a variety of grid environments.

ProActive

<http://www-sop.inria.fr/oasis/proactive/>

ProActive is the Java GRID middleware library (with Open Source code under LGPL license) for parallel, distributed and multi-threaded computing. With a reduced set of simple primitives, ProActive provides a comprehensive API to simplify the programming of Grid Computing applications: distributed on Local Area Network, on clusters of workstations, or on Internet Grids.

Security Projects

To protect the vital infrastructure and information, security is a constant evolving requirement of grid computing. These projects represent some of the cutting edge new security standards and implementations of grid security solutions.

GridShib

<http://gridshib.globus.org/>

GridShib: Integrating federated authorization infrastructure (Shibboleth) with Grid technology (the Globus Toolkit) to provide attribute-based authorization for distributed scientific communities.

GUMS

<http://grid.racf.bnl.gov/GUMS/>

Grid User Management System (GUMS) is a Grid Identity Mapping Service. Identity mapping is necessary when a site's resources do not use GRID credentials natively, but instead use a different mechanism to identify users, such as UNIX accounts or Kerberos principals.

PRIMA

<http://computing.fnal.gov/docs/products/voprivilege/prima/prima.html>

PRIVilege Management and Authorization (PRIMA) is a system which provides enhanced grid security. PRIMA is both a comprehensive grid security model and system. In PRIMA, a privilege is a platform independent, self-contained representation of a fine-grained right. PRIMA achieves platform independence of privileges by externalizing fine-grained access rights to resource objects from the resource's internal representation.

Resource Management & Scheduling

An essential component of grids are to manage and schedule jobs across resources. These projects demonstrate a few different strategies.

Community Scheduling Framework

http://www.globus.org/grid_software/computation/csf.php

Community Scheduler Framework (CSF) is an open source implementation of an OGSA-based meta-scheduler. It supports the emerging WS-Agreement specification and the Globus Toolkit's GRAM service. CSF fills in gaps in the existing resource management picture and it is integrated with Platform LSF and Platform Multicluster. The CSF open source project is included in the Globus Toolkit 4.0 release.

SPRUCE

<http://spruce.teragrid.org>

High-performance modeling and simulation are playing a driving role in decision making and prediction. For time-critical emergency support applications such as severe weather prediction, flood modeling, and influenza modeling, late results can be useless. A specialized infrastructure is needed to provide computing resources quickly, automatically, and reliably. SPRUCE, the Special PRiority and Urgent Computing Environment is a system to support urgent or event-driven computing on both traditional supercomputers and distributed Grids.

Grid Resource Monitoring

Monitoring resources and applications is key to the success of grids. Through an easy to use interface, these sophisticated tools help users gather, catalog, and monitor various types of resources. Moreover, systems administrators are also able to monitor the health of their grids. These evolving grid projects list a few of the open source options:

GridCat

<http://www.ivdgl.org/gridcat/home/index.html>

GridCat is a high level grid cataloging system using status dots on geographic maps as well as catalog. The maps help debug site troubles. The catalog contains information on site readiness with many other valuable information per site to help job submission and job scheduling for application users and grid scheduler developers. GridCat tries to present the grid-site at its simplest status representation.

Gridscape II

<http://www.gridbus.org/gridscape/>

Gridscape II, a customized portal component that can be used on its own or plugged in to compliment existing Grid portals. Gridscape II manages the gathering of information from arbitrary, heterogeneous and distributed sources and presents them together seamlessly within a single interface. It leverages the Google Maps API in order to provide a highly interactive user interface. Gridscape II is simple and easy to use, providing a solution to those users who don't wish to invest heavily in developing their own monitoring portal from scratch, and also for those users who want something that is easy to customize and extend for their specific needs.

Storage & Data Management

From open source, high-performance file systems to seamless access of data from heterogeneous environments, the following projects bring together and optimize a variety of storage and data management solutions. This track emphasizes storing, managing, and moving data across resources and connecting data resources over a network.

LUSTRE

<http://www.lustre.org/>

The Lustre File System, an open source, high-performance file system from Cluster File Systems, Inc., is a distributed file system that eliminates the performance, availability, and scalability problems that are present in many traditional distributed file systems. Lustre is a highly modular next generation storage architecture that combines established, open standards, the Linux operating system, and innovative protocols into a reliable, network-neutral data storage and retrieval solution. Providing high I/O throughput in clusters and shared-data environments, Lustre also provides independence from the location of data on the physical storage, protection from single points of failure, and fast recovery from cluster reconfiguration and server or network outages.

NeST: Network Storage

<http://www.cs.wisc.edu/condor/nest/>

NeST is a software network storage device providing secured storage allocation for a specific time period. The size and duration of allocation units or lots are negotiated between NeST and the user or application. These lots can also be expanded in size, extended in time and/or subdivided into a hierarchy. Plus NeST offers access control lists for both lot and file access. NeST offers multiple protocol interfaces, including its internal 'Chirp', HTTP and GSI-FTP.

SAMGrid

<http://projects.fnal.gov/samgrid/>

SAMGrid is a general data handling system designed to be a key device for experiments with large (petabyte-sized) datasets and widely distributed production and analysis facilities. The components now in production provide a versatile set of services for data transfer, data storage, and process bookkeeping on distributed systems.

UberFTP

<http://dims.ncsa.uiuc.edu/set/uberftp/index.html>

Building upon the technologies of GridFTP, UberFTP is the first interactive, GridFTP-enabled ftp client. The basic GridFTP client is not interactive and allows only one file transfer at a time. UberFTP provides an interactive tools that works much like the popular NCFTP tool. It supports GSI authentication, parallel data channels and third party transfers.

Conclusion

Grid computing is one of the most exciting technologies that are having powerful effects on the way we solve complex problems and share diverse resources. In addition to cancer and physics, it also has great influence on security and authentication, discovery, monitoring, and information services, data management, resource management and scheduling. This article was developed to give you an overview into the important production projects taking place and I am sure we're in for more.

This article will updated quarterly so please check back again for the latest resources.

About the author

Edna Nerona is the owner of Legacy Studios, Inc, a creative services firm located in San Diego, CA. She has BA in Journalism from San Diego State University and has previously worked for the San Diego Supercomputer Center and Entropia, Inc. Edna is also an accomplished speaker with Toastmasters International.