Grid Developers: Open Source Tools -- Starter Set

Level: Introductory

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

01 Dec 2006

Learn more about Grid computing deployment with this reading list compiled for new users, programmers and administrators.

Introduction

Since you've gotten familiar with the fundamentals and exciting benefits of grid computing, we'd now like to guide you through the process of getting your grid project started and deployed. This list of recommended reading material will provide a foundation for building and getting your grid project off the ground. Taken from a variety of online sources, these resources have been selected with the intention of introducing programmers and administrators plus new users looking for specific information about using, deploying and developing on grid infrastructure.

As you previously read, grid computing pools and maximizes the value of computing resources. As such, this list describes resources that apply to learning about how to deploy and administer your own grid, how to develop grid computing applications.

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

Getting started

This track provides resources you'll need to get started with grid development.

Deployment

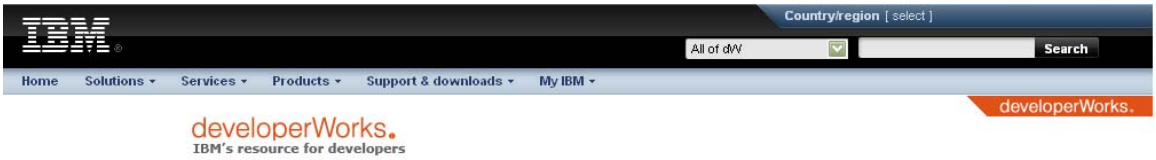
Creating or deploying applications for a grid requires a reliable grid deployment environment to test and develop with. A good place to start building your own grid is with the Globus Toolkit.

Globus Toolkit
http://globus.org/toolkit/

To build a grid today, a good place to start is to download the Globus Toolkit. Developed by The Globus Project, a research and development project that focuses on enabling the application of grid concepts to scientific and engineering computing, the open source tool kit is a set of services and software libraries designed to support grids and grid applications.

Introduction to Grid Computing with Globus
http://www.redbooks.ibm.com/abstracts/sg246895.html

This IBM Redbook offers readers a discussion of the first basics of grid computing, in addition to installation and configuration information for building your own grid using Globus.



Comprehensive Toolkits

Installing grid software is challenging and time consuming. The following focuses on "open source" toolkits, which are readily available and make it easier for you to build, deploy and manage your grid project. Some are packaged as a set of components that can be used either independently or in concert to develop applications.

Grid Infrastructure Toolkits -

Below is a list of "open source" toolkits that offer the entire grid infrastructure to get you started on building your own grid:

Globus

http://www.globus.org/toolkit/
The Globus® Toolkit is an open source software toolkit used for building grids.

The Virtual Data Toolkit (VDT)

http://vdt.cs.wisc.edu/
The goal of the VDT is to make it as easy as possible for users to deploy, maintain and use grid middleware.

Application Development Toolkits

Below is a list of toolkits to help develop or port your existing applications to a grid:

Java Commodity Grid Toolkit

http://wiki.cogkit.org/index.php/Java_CoG_Kit
The Java Commodity Grid (Java CoG) Kits allow Grid users, Grid application developers, and Grid administrators to use, program, and administer Grids from a higher-level framework. The Java CoG Kit jGlobus module provides the basic APIs to the Grid to allow access to gridFTP servers, the classic GRAM services, and a complete implementation of GSI.

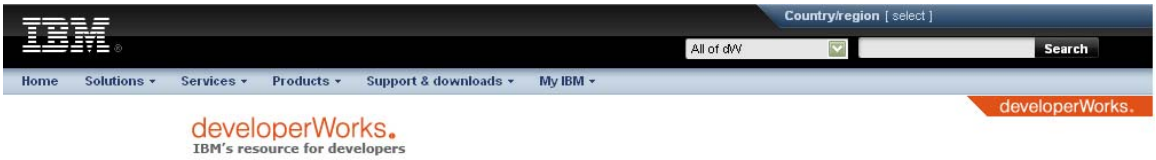
PyGlobus/Python Commodity Grid Toolkit

http://wiki.cogkit.org/index.php/Python_CoG_Kit
The Python CoG Kit is also known under the name pyGlobus. pyGlobus uses SWIG to generate Python interfaces to Globus.

Grid Application Toolkit

http://www.gridlab.org/WorkPackages/wp-1/
GAT is a set of coordinated, generic and flexible APIs for accessing Grid services from generic application codes, portals, data managements systems, together with working implementations provided by the tools developed in the Grid Lab project. GAT is designed in a modular plug-and-play manner, such that tools developed anywhere can be plugged into GAT.

WSRF for WebSphere Application Server



http://www.alphaworks.ibm.com/tech/wsrf4was
Provides support for the implementation of Web Service Resource Framework (WS-RF) for WebSphere Application Server. It introduces a Java interface for WS-Addressing EndpointReferences (EPR) and exploits the capabilities of WS-Addressing to support the WS-Resource access pattern described in the OASIS WSRF Technical Committee's draft specification for WS-Resources.

Grid Application Framework for Java

http://www.alphaworks.ibm.com/tech/GAF4J
A framework that abstracts all grid semantics from the application logic and provides a simpler programming model that lines up smoothly with common Java programming models.

GridFTP.NET and Gram.NET

http://www.cs.virginia.edu/~gsw2c/GridTools/GridTools.htm
This site provides client and server implementations of two of the most popular services provided by the Globus Toolkit. The Globus Toolkit v. 4 (GT4) implementation of GridFTP, a data transfer protocol, and GRAM, a job execution protocol, are widely deployed in many scientific grids. We present implementations of both GridFTP and GRAM, services and clients, that run on the Microsoft .NET Framework and are interoperable with their GT4 counterparts.

Portal Toolkits

Web portals are gateways to grids that enable easy access to for users to access a grid through the web browser. Here is a list of portal toolkits to help develop your own grid portal:

Open Grid Computing Environments (OGCE) Portal Software

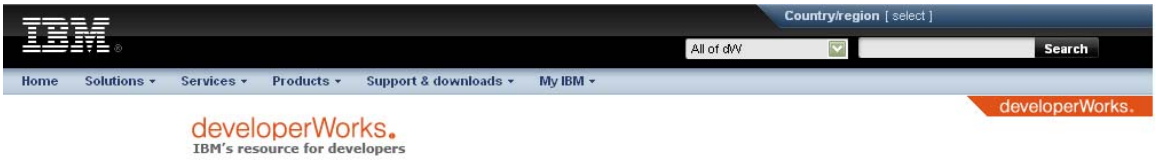
http://www.collab-ogce.org/ogce2/
The OGCE collaboration develops JSR 168-compatible portlets and Web services for building Web portals for science gateways. The OGCE release combines into a single download package everything you need to run and develop your own Grid portal.

GridSphere Portal Framework and GridPortlets

http://www.gridsphere.org
The GridSphere portal framework provides an open-source portlet based Web portal. GridSphere enables developers to quickly develop and package third-party portlet web applications that can be run and administered within the GridSphere portlet container. The Grid Portlets web application puts the "Grid" in GridSphere. Grid Portlets builds upon the core features in the GridSphere portal framework to provide developers with a framework for developing Grid portals.

Security & Authentication

Grid computing is unique because it requires coordinating diverse resources. Security is a crucial component in grid computing and applications must work with existing security infrastructure



available on grids. The following resources cover grid security:

Grid Security Infrastructure (GSI)

http://www.globus.org/toolkit/security/

Security tools are concerned with establishing the identity of users or services (authentication), protecting communications, and determining who is allowed to perform what actions (authorization), as well as with supporting functions such as managing user credentials and maintaining group membership information. These components are all essential components that are part of GSI. Make sure to carefully learn and understand these pieces since they are crucial to interact and develop applications on the grid.

MyProxy

http://grid.ncsa.uiuc.edu/myproxy/

MyProxy is a credential repository for the Grid. Storing your Grid credentials in a MyProxy repository allows you to retrieve a proxy credential whenever and wherever you need one, without worrying about managing private key and certificate files. Using a standard web browser, you can connect to a Grid portal and allow the portal to retrieve a proxy credential for you to access Grid resources on your behalf. MyProxy servers are used in many production grids and grid applications, check it out and see how it can help your application development.

Community Authorization Service (CAS)

http://www-unix.globus.org/toolkit/docs/4.0/security/cas/

Managing individual accounts for each user on your grid can be cumbersome and doesn't always make sense with your application. What if you want to enable access to a group of users? One option is the Community Authorization Service. Building on GSI, the CAS component of the Globus Toolkit allows resource providers to specify coarse-grained access control policies in terms of communities as a whole, delegating fine-grained access control policy management to the community itself.

VOMS: Virtual Organization Membership Service

http://edg-wp2.web.cern.ch/edg-wp2/security/voms/voms.html

VOMS is a system for managing authorization data within multi-institutional collaborations. VOMS provides a database of user roles and capabilities and a set of tools for accessing and manipulating the database and using the database contents to generate Grid credentials for users when needed.

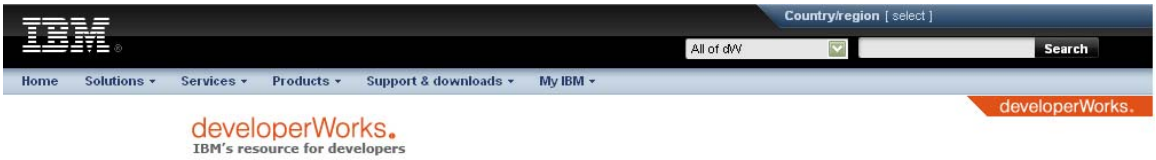
Discovery, Monitoring, & Information Services

An important aspect of grids is being able to monitor and discover new systems on the grid. This includes gathering information about services and resources. This collection of services offers a variety of options:

Globus Monitoring and Discovery System (MDS)

http://www.globus.org/toolkit/mds/

The Monitoring and Discovery System (MDS) is the information services component of the Globus Toolkit and provides information about the available resources on the Grid and their status.



Monitoring Agents using a Large Integrated Services Architecture (MonALISA)

http://monalisa.cacr.caltech.edu

The MonALISA system is designed as an ensemble of autonomous multi-threaded, self-describing agent-based subsystems which are registered as dynamic services, and are able to collaborate and cooperate in performing a wide range of information gathering and processing tasks. These agents can analyze and process the information, in a distributed way, to provide optimization decisions in large scale distributed applications.

Ganglia

http://ganglia.sourceforge.net/

Ganglia is a scalable distributed monitoring system for high-performance computing systems such as clusters and Grids.

Inca Test Harness and Reporting Framework

http://inca.sdsc.edu/

Inca is a flexible framework for the automated testing, benchmarking and monitoring of Grid systems. It includes mechanisms to schedule the execution of information gathering scripts and to collect, archive, publish, and display data.

Data Management

One of the appealing traits of grid computing is that it can bring variety of resources together. The following applications provide a means of storing, managing, and moving data across resources and as well as connecting data resources over a network.

Globus GridFTP

http://www.globus.org/grid_software/data/gridftp.php

GridFTP is a high-performance, secure, reliable data transfer protocol optimized for high-bandwidth wide-area networks. It is based upon the Internet FTP protocol, and it implements extensions for high-performance operation.

Storage Resource Broker (SRB)

http://www.sdsc.edu/srb

Storage Resource Broker (SRB) supports shared collections that can be distributed across multiple organizations and heterogeneous storage systems. The SRB can be used as a Data Grid Management System (DGMS) that provides a hierarchical logical namespace to manage the organization of data (usually files).

Resource Management & Scheduling

A crucial aspect of grids is being able to manage resources and schedule jobs across a grid. This can be done through various resource management systems:

Globus GRAM

http://www.globus.org/toolkit/gram/

The Grid Resource Allocation and Management (GRAM) service provides a single interface for requesting and using remote system resources for the execution of "jobs". The most common use of GRAM is remote job submission and control. It is designed to provide a uniform, flexible interface to job scheduling systems.

Condor

http://www.cs.wisc.edu/condor/

Open source workload management software which matches resource owners with resource consumers. Designed to run a job from a single administrative domain. Includes the ability to submit jobs through Globus across many administrative domains.

Conclusion

Years ago, no one could imagine the internet, a technology that allows users everywhere to share information. Now we are approaching another exciting time where we can share information as well as applications and computing power. Grid computing has many revolutionary technologies. We created this article to help you become more familiar with infrastructure, deployment, security, and data management.

In addition to this article, IBM has a list of open source and licensed software tools and technologies on the Grid Computing Downloads & Products Page (http://www-128.ibm.com/developerworks/views/grid/downloads.jsp)

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.