"The Availability Corner"
Advice and Solutions for Enterprise Computing

As Seen in *The Connection*, An ITUG Publication
November 2004 – September 2006

About the Authors:

Dr. Bill Highleyman, Paul J. Holenstein, and Dr. Bruce Holenstein, have a combined experience of over 90 years in the implementation of fault-tolerant, highly available computing systems. This experience ranges from the early days of custom redundant systems to today's fault-tolerant offerings from HP (NonStop) and Stratus.

Series Topics:

Gravic, Inc.
Shadowbase Products Group
17 General Warren Blvd.
Malvern, PA 19355
610-647-6250
www.ShadowbaseSoftware.com

The Availability Corner

Grid Computing

March/April, 2006

Dr. Bill Highleyman
Dr. Bruce Holenstein
Paul J. Holenstein

In many organizations, massive data centers are dedicated to managing an organization's computing, transaction management, and data management requirements using systems dedicated to certain tasks. Sometimes, some systems are underutilized while others struggle with capacity. At other times, the roles reverse.

Wouldn't it be nice if these computing resources could become non-dedicated? Then they could be assigned to the tasks at hand, significantly reducing the amount of required hardware and its consequent management. Moreover, should a resource fail, there would be others to take over. This is HP's vision of a computing utility. Users only use the capacity that they require at any one time, and all of the data center capacity is available to all the users.

This concept has led to an industry movement called Grid Computing, which seeks to standardize the use of disparate systems in a computing utility. The utility would manage all of the resources available to it as a set of services to improve reliability, availability, business agility, security, and a better return on IT investment.

Grid Computing has actually been around for a long time in different forms. For years, government and academic systems have been linked globally to create, in essence, supercomputers contributing to advances in medicine, physics, meteorology, and other fields. The SETI@home project allows home PC users to contribute time on their systems toward the search for extraterrestrial life.

Another notable example dating back to the 1970s is the HP NonStop server (originally from Tandem Computers). This system manages the assignment of various application workloads in a multi-processor system, adding or removing processors assigned to a specific application as its demands warrant. In fact, the NonStop server was used as a prototype for the grid.[1]

Grid computing has become a major industry initiative. It is important to note that it is not an off-the-shelf product. Rather, it is a set of standards. The standards-setting body for grid computing is the Global Grid Forum (GGF). The GGF was founded in 1999 by HP, IBM, Sun, Microsoft, Intel, and SGI, among others.

---

[1] See Andrew Chien's article in "*The Grid2: Blueprint for a New Computing Architecture,"* a collection of articles edited by Ian Foster and Carl Kesselman, Morgan Kaufmann; 2004.

The fundamental specification for the grid architecture is the Open Grid Services Infrastructure (OGSI). This specification and a variety of associated papers from grid primers to detailed white papers can be found at www.ggf.org/ggf_docs_final.htm.

The OGSI (which is actually a working group within GGF) focuses on the mechanisms that enable resources to be discovered, managed, and exploited on a large scale by defining an Open Grid Services Architecture (OGSA). The heart of the grid is the resources available to it – the applications, the databases, and the physical resources such as the processing systems. Each resource registers with a Registry (a directory) exposing its capabilities.

When a user wishes to acquire resources, it finds them in the Registry and then binds to those resources. At this point, the user has acquired the resources it needs and can proceed with its work. To aid in the effective allocation of resources, a user can reserve resources in advance.

The services described by OGSI include resource scheduling, workload balancing, standardized interfaces, authentication and access control, performance monitoring, and problem diagnosis.

The OSGI standard is based on Web Services as defined by the Web Services Definition Language (WSDL), with extensions. Thus, a grid implementation may be built on existing systems for hosting web services such as J2EE.

The primary implementation of grid services per the OGSI specification is Globus, a toolkit created by the Globus Consortium, founded by HP, IBM, Sun, and Intel. This consortium includes many other industry and government organizations. The Globus Consortium is focused on the advancements of the commercial use of the Globus software.

At the current time, the grid is still in the early stages of specification and acceptance. Almost all industry vendors have announced that they do, or will, support the grid standards. The beginning of a comprehensive grid toolkit, Globus, is available. But full-fledged commercial grid applications have yet to be seen (marketing aside).

Grid Computing is an exciting concept whose time has perhaps come. However, will this industry initiative suffer the fate of other strongly supported initiatives such as DCE (Distributed Computing Environment)? DCE was a standard for allowing disparate processing systems to interoperate. However, it never seemed to gain wide-spread acceptance in the industry. Hopefully Grid Computing will not suffer this fate. But for the timid or conservative, pay heed to the old adage: “Caution: Standards may be hazardous to your health.”