Database Integration in the Grid Infrastructure

Emmanuel Udoh
Indiana University – Purdue University, USA

INTRODUCTION

The capability of the Web to link information over the Internet has popularized computer science to the public. But it is the grid that will enable the public to exploit data storage and computer power over the Internet analogous to the electric power utility (a ubiquitous commodity). The grid is considered the fifth generation computing architecture after client-server and multitier (Kusnetzky & Olofson, 2004) that integrates resources (computers, networks, data archives, instruments, etc.) in an interoperable virtual environment (Berman, Fox, & Hey, 2003). In this vein, grid computing is a new IT infrastructure that allows modular hardware and software to be deployed collectively to solve a problem or rejoined on demand to meet changing needs of a user.

Grid computing is becoming popular in the enterprise world after its origin in the academic and research communities (e.g., SETI@home), where it was successfully used to share resources, store data in petabytes/exabytes, and ultimately lower costs. There are several reasons for the embrace of the enterprise grids. In the nineties, the IT world was confronted with the high cost of maintaining smaller, cheaper and dedicated servers such as UNIX and Windows. According to Oracle (2005), there was the problem of application silos that lead to underutilized hardware resources; monolithic and unwieldy systems that are expensive to maintain and change; and fragmented and disintegrated information that cannot be fully exploited by the enterprise as a whole. Various surveys put the average utilization of servers in a typical enterprise to often much less than 20% (Goyal & Lawande, 2006; Murch, 2004). But with the increasingly available cheaper, faster and affordable hardware such as server blades, and operating systems like the open source Linux, the IT world embraced grid computing to save money on hardware and software. With the growing importance of grid computing, it is easy to conjecture why many new terms have been coined for it. In the literature and industry, other terms used interchangeably for grid computing are utility computing, computing on demand, N1, hosted computing, adaptive computing, organic computing and ubiquitous computing (Goyal & Lawande, 2006; Murch, 2004; Oracle, 2005).

The grid is an all-encompassing, 21st century computing infrastructure (Foster, 2003; Joseph & Fellenstein, 2004) that integrates several areas of computer science and engineering. A database is an important component of the application stack in the industry and is increasingly being embedded in the grid infrastructure. This article focuses on integration of database grids or grid-accessible databases in the industry using Oracle products as examples. Vendors like Oracle and IBM are providing grid-enabled databases that are supposed to make enterprise systems unbreakable and highly available. Oracle has been in the forefront in this endeavor with its database products. In recognition of the significant new capabilities required to power grid computing, Oracle has named its new technology products Oracle 10g (g for grid). Oracle provides seamless availability through its database products with such features like streams, transportable tablespaces, data hubs, ultra-search and real application clusters. Although companies will not like to distribute resources randomly on the Internet, they will embrace enterprise database grids, as they embraced Internet in the form of Intranets. To the business world, database grids will help achieve high hardware utilization and resource sharing, high availability, flexibility, incrementally scalable low cost components and reduced administrative overhead (Kumar & Burleson, 2005; Kusnetzky & Olofson, 2004).

BACKGROUND

The grid technology is still evolving, and databases are increasingly being incorporated into its infrastructure. IBM has contributed immensely in this endeavor with its autonomic and grid computing projects, but Oracle is the clear-cut industry leader in enterprise database grids. Oracle, a member of the enterprise grid alliance (EGA) that promotes tools and standards for enterprise computing, has been preparing and targeting the grid market since the Oracle9i release with products like Oracle real application clusters (now morphed to automatic storage management system). This article will therefore discuss enterprise grid computing in terms of the features available in Oracle 10g database products.

The main features that clearly differentiate grid computing from other forms of computing architectures, such as client server or multitier, are virtualization and provisioning. Virtualization creates a logical view or abstraction that allows the pooling together of resources (e.g., data, computing power, storage capacity, and other resources) for consumption by other applications, while provisioning determines how to meet on demand the specific needs of consumers.

As consumers request resources through the virtualization layer (which breaks the hard-coded connection between providers and consumers (Oracle, 2005), provisioning guarantees resources are allocated for the request. To achieve these objectives, a grid implements a layered architecture as depicted in Figure 1a.

In a semblance of the generalized grid architecture, the Oracle grid builds a stack of its software in a virtual environment, as shown in Figure 1b. The bottom layer hosts the storage units such as a storage area network (SAN), while the next horizontal layer contains the infrastructure such as the hardware and software that create a data storage and program execution environment (infrastructure grid). The next layer, the application server, contains the program logic and flow that define specific business processes (application grid). The topmost layer (information grid) hosts applications such as user applications, enterprise resource planning and portal software that can be accessed over the network without the application being architected to support the device or network. This virtualized environment has a unified management structure as well as an infrastructure for security.

As depicted in Figure 1b, Oracle 10g software stack, which is configured to self-manage, acts as a single computing resource to the user even in a geographically distributed environment. This allows organizations to protect and optimize their investment in hardware and software and also access newer system features in a more reliable, powerful and scalable environment. To keep this unified structure manageable and also eliminate the application silo model, Oracle enterprise manager has a grid control that monitors, provisions, clones and automates even in heterogeneous environments (Khilani, 2005).

MAIN FOCUS

The evolving enterprise database grid brings substantial benefits to the industry but poses a major challenge in integration and adoption. With yearly global business volume in excess of $10 billion dollars, databases are critical components in the enterprise application stack. The existing database infrastructure in companies are aggregates of many years of investments in a wide range of interfaces and tools for performance and security. Noting all the hype about grids, it is natural that its adoption will be resisted by employees, who are unwilling to change existing processes and technology. Researchers believe that building grid-enabled database infrastructures from scratch is both unrealistic and a waste of effort and resources (Watson, 2003). Instead, existing database management systems should be integrated into the grid in an incremental manner, absorbing the existing investments without being disruptive. Oracle's grid-enabled database takes cognizance of this reality, and hence this article focuses on database grid integration and adoption vis-a-vis Oracle grid products.

At the department of computer science (CS), Indiana University–Purdue University, Fort Wayne, an Oracle database grid was introduced to aid the administration of the database program (for research and education). Experience from this program supports these transitioning steps for the integration and adoption of an enterprise database grid: identification, standardization, consolidation and automation.

Identification

Organizations have different IT infrastructures that may influence decisions to integrate their enterprises in database grids.

Figure 1. Structure of generalized grid and Oracle database grid
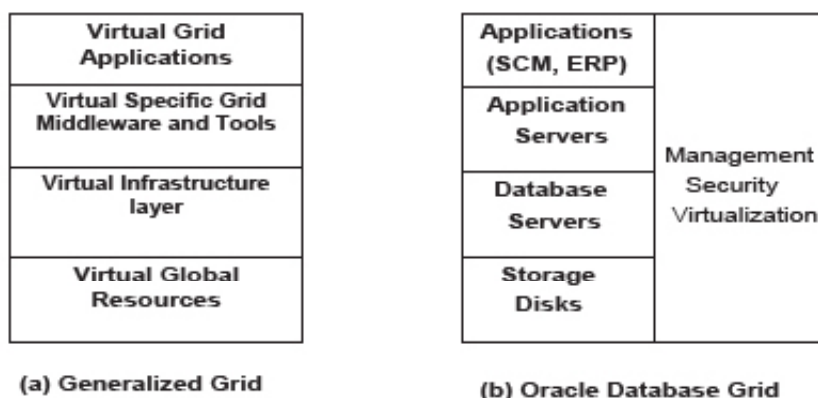
(a) Generalized Grid

(b) Oracle Database Grid

Related Content

Social Issues in Electronic Commerce: Implications for Policy Makers
Anastasia Papazafeiropoulouand Athanasia Pouloudi (2001). *Information Resources Management Journal (pp. 24-32).*
www.irma-international.org/article/social-issues-electronic-commerce/1190

Knowledge Management for Development (KM4D)
Alexander G. Flor (2019). *Advanced Methodologies and Technologies in Library Science, Information Management, and Scholarly Inquiry (pp. 392-401).*
www.irma-international.org/chapter/knowledge-management-for-development-km4d/215941

Fault Tolerance for Distributed and Networked Systems
Wenbing Zhao, Louise E. Moserand P. Michael Melliar-Smith (2005). *Encyclopedia of Information Science and Technology, First Edition (pp. 1190-1196).*
www.irma-international.org/chapter/fault-tolerance-distributed-networked-systems/14409

Classification of Printed Moroccan Town and Village Names
Said Nouriand Mohamed Fakir (2014). *Journal of Information Technology Research (pp. 1-11).*
www.irma-international.org/article/classification-of-printed-moroccan-town-and-village-names/124910

Survey of Breast Cancer Detection Using Machine Learning Techniques in Big Data
Madhuri Guptaand Bharat Gupta (2019). *Journal of Cases on Information Technology (pp. 80-92).*
www.irma-international.org/article/survey-of-breast-cancer-detection-using-machine-learning-techniques-in-big-data/227680