

Chapter 6.3

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 table-spaces, 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.

6 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/database-integration-grid-infrastructure/8012

Related Content

Data Warehouse Design to Support Customer Relationship Management Analysis

Colleen Cunningham, Il-Yeol Song and Peter P. Chen (2006). *Journal of Database Management* (pp. 62-84).

www.irma-international.org/article/data-warehouse-design-support-customer/3353

A Scalable Algorithm for One-to-One, Onto, and Partial Schema Matching with Uninterpreted Column Names and Column Values

Boris Rabinovich and Mark Last (2014). *Journal of Database Management* (pp. 1-16).

www.irma-international.org/article/a-scalable-algorithm-for-one-to-one-onto-and-partial-schema-matching-with-uninterpreted-column-names-and-column-values/138623

Privacy-Preserving Data Mining

Alexandre Evfimievski and Tyrone Grandison (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 527-536).

www.irma-international.org/chapter/privacy-preserving-data-mining/20737

Mediating RDF/S Queries to Relational and XML Sources

I. Koffina, G. Serfiotis, V. Christophides and V. Tannen (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 596-614).

www.irma-international.org/chapter/mediating-rdf-queries-relational-xml/7933

Innovative Technological Paradigms for Corporate Offshoring

Tapasya Patki and A. B. Patki (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 1490-1510).

www.irma-international.org/chapter/innovative-technological-paradigms-corporate-offshoring/7988