Virtualization of Enterprise Data and Services

George Feuerlicht and Emine Ates

Faculty of Information Technology,
University of Technology, Sydney,
P.O. Box 123, Broadway, 2007, NSW, Australia
jiri@it.uts.edu.au

Abstract

Present IT environment in most organizations is too complex and highly inefficient in resources utilization. Many experts agree that the current generation of enterprise information systems configured using dedicated resources to support individual applications and business processes leads to significant underutilization of resources and high ongoing support costs. It is becoming clear that in order to improve return on IT investment enterprise IT environment needs to be dramatically simplified, integrated and consolidated. This necessitates revision of the entire approach to enterprise computing and development of a radically new IT infrastructure based on the Utility Computing model. IT resources such as storage, server capacity, database and application services will be delivered on demand from data centers using dynamically configurable computing grids. The realization of the Utility Computing vision relies on achieving virtualization of IT resources, so that storage, servers, and other resources can be dynamically configured to provision application services as need arises. In this paper we first discuss today's enterprise computing challenges, in particular the high costs associated with deploying and operating enterprise applications. We then discuss how these challenges can be addressed using consolidation and virtualization of enterprise data and services.

Keywords

Resource Virtualization, IT consolidation, Data Integration, Grid Computing

1 Introduction

Notwithstanding major advances over the last three decades, enterprise computing is regarded by many as inflexible, very expensive and associated with poor return on investment. Successive generations of enterprise computing architectures and approaches have attempted to address computing requirements of user organizations with most up-to-date hardware and software technology platforms, each time claiming substantial benefits over previous approaches. Starting with centralized, mainframe solutions in the 1970s, progressing to client/server systems in 1980s, and Internet computing applications in 1990s organizations have attempted to support their increasingly complex business processes and at the same time to control the cost of IT solutions.

Centralized mainframe computing was associated with poor scalability and limited flexibility, and generally regarded as not being able to keep up with user demands for applications. The main attraction of client/server systems was improved user interface and scalability achieved by distributing the processing load across multiple server and client platforms, with some configurations supporting thousands of concurrent users. However, this lead to the proliferation of client (PC) workstations and various types of servers (e.g. database servers, application servers, etc), resulting in an environment where the ongoing maintenance and administration costs constitute a major component of the Total Cost of Ownership (TCO). The emergence of three-tier client/server systems, while enhancing scalability has introduced another layer of complexity with additional middleware servers. This situation was further exacerbated by the advent of Internet computing with massive increase in the size of user populations and corresponding increase in the number, size and complexity of the server platforms. Unlike centralized mainframe systems that are characterized by limited capacity to accommodate users and applications, distributed systems in general and client/server systems in particular typically provide excess processing capacity and suffer from resource underutilization. Modern enterprise computing environments typically involve independently developed and deployed applications systems (e.g. ERP applications) each configured to accommodate the maximum processing load and storage. As individual application systems are unable to share resources effectively, most server and storage resources remain significantly underutilized and in some cases idle at least some of time. In retrospect, distribution of computing resources associated with client/server and Internet computing models, while achieving good scalability and many other benefits, also produced a situation where most resources are poorly utilized, difficult to administer and expensive to maintain. Many large global organizations are struggling to control the costs associated with administering, maintaining and upgrading hundreds, and in some cases thousands of servers scattered in different locations around the world.

A new enterprise computing paradigm is emerging to address these problems, based on the concept of Utility Computing where *re-centralized* and shared computing resources are managed together in a data centre and deployed on-demand, i.e. as needed by individual enterprise applications. Closely related to the Utility Computing model is the idea of application services, i.e. enterprise applications delivered as services over the Internet. While the Application Service Provider (ASP) model is not new, the use of advanced Grid Computing architectures will enable delivery of highly scalable application services in a reliable manner to large user populations. These technological advances distinguish Utility Computing from the earlier ASP and outsourcing models, and will ultimately result in significant reduction in the costs of enterprise software solutions. However, achieving the vision of Utility Computing in the context of enterprise computing presents a number of important challenges. A key requirement is the virtualization of computing resources, so that storage, servers, and other resources can be dynamically configured to provision enterprise-level application services on-demand. Effective virtualization of computing resources is not achievable in an environment characterized by excessive heterogeneity and general lack of standardization. Consolidation and standardization of enterprise computing resources is the first step towards resource virtualization, and ultimately towards Utility Computing.

In this paper we first discuss consolidation of computing resources (section 2) and then describe approaches to resource virtualization (section 3), and Grid Computing (section 4). In conclusion (section 5) we summarize the benefits and challenges of Utility Computing.

2 Consolidation of Enterprise Computing Resources

Current enterprise computing environments require administration of a large number of resources distributed around the organization, each dedicated to a specific application. Moreover enterprise applications also need development and backup/standby resources (e.g. databases) to improve availability and support change management procedures. In such environments average resource utilization is low, usually around 30%. Optimizing performance of individual applications is not a cost-effective proposition as additional investment is needed for each *application silo* [Yuhanna, 2002]. To overcome such issues, and to improve overall resource utilization, many organizations are embarking on a process of consolidation of IT resources with the aim of reducing the complexity of their computing environments.

The main idea of consolidation is to reduce the number of IT resources (e.g. servers) to a minimum level without sacrificing functionality and to minimize heterogeneity of the environment by standardizing server, network, storage and DBMS hardware and software, resulting in the reduction of overall operational costs [Yuhanna, 2003]. Organization-wide consolidation of IT resources involves consolidating resources such as servers, databases and storage into data centers in order to simplify the management of the overall IT infrastructure [Quest, 2002]. Some organizations are building large data centers and use clustering and grid technologies to ensure scalability and reliability of hosted enterprise applications. Consolidation has many benefits including improved security, availability, reliability and performance, and simplified centralized management and maintenance. However consolidation is not a trivial task and requires careful planning and implementation. A prerequisite for successful consolidation is the standardization of technology platforms at various levels of the enterprise IT architecture, including. network protocols, hardware platforms, operating systems, DBMS platforms and applications [Dell, 2004], [META, 2004]. Some large organizations are implementing multiple geographically distributed data centers to take advantage of different time-zones and reduced costs associated with individual countries and geographic regions.

2.1. Storage Consolidation

Organizations today need to manage massive amounts of various types of data and typically require substantial investment in data storage and data management solutions. Storage consolidation aims to increase efficiency of these data services with improving overall resource utilization and storage management [Subbiah, 2004]. Storage consolidation involves de-coupling storage resources from individual servers and deploying NAS (Network Attached Storage) or SAN (Storage Area Network) technologies to provide shared storage for all applications. Storage consolidation improves the overall storage utilization but most of the savings are usually associated with reduction in operational and maintenance costs. NAS typically provides more efficient file-level access reducing TCO but is not optimized for database access. SAN pools storage capacity providing more efficient and scalable storage solution. Also, since SAN uses its own dedicated network, it enables faster, more flexible and reliable data access; this makes SAN a popular storage consolidation technology, in particular for large organizations. In general, NAS storage is used as a front-end technology to share files and SAN as a back-end technology to share resources on common networks [Krantz, Ryan, 2004], [Subbiah, 2004].

2.2 Server Consolidation

Server consolidation aims to address the problem of management of large number of servers dedicated to different enterprise applications. Typically, different kinds of servers, e.g. application and database servers can be consolidated into a relatively small number of powerful clustered servers that are easier to manage. Technologies such as blade servers can be used to distribute workload uniformly between servers, in a manner transparent to the applications. This provides more efficient resource sharing and increased availability and performance. However, in such consolidated environment the requirements of every application are considered uniform and this can make optimization of individual application rather difficult [Quest, 2002]

2.3 Database Consolidation

Database consolidation is a further step towards simplification of management of IT resources. Database consolidation involves reduction of the number of database instances and migrating databases to centralized database servers. Database objects originally stored in separate databases are migrated to a centralized database to enable sharing of resources such as storage and processors. Database cluster technology is used to provide parallel access to the database from multiple cluster nodes. Shared nothing or shared disk clustering technologies can be implemented according to application requirements and data sharing needs [Subbiah, 2004], [Yuhanna, 2002]. Database consolidation reduces TCO, and in the long term increases ROI. Major challenges of database consolidation are high initial costs and *political* issues associated with data ownership. Individual departments that control and manage their own databases may be reluctant to hand over control to a centralized authority; issues of data ownership and control need to be carefully addressed during the process of database consolidation. Another challenge is to predict future growth and workload to ensure optimal performance in the long term. Most leading database vendors (e.g. Oracle, IBM and Microsoft) already provide some level of support for database consolidation. Database consolidation is a challenging exercise, in particular if issues of data heterogeneity are involved, and needs to be implemented step by step, progressively demonstrating the benefits of consolidation [Kirby, 2004].

3 Resource Virtualization

As noted above virtualization is a key requirement for improving the utilization of computing resources. Virtualization services are supported by a range of technological solutions including storage technologies (i.e. NAS and SAN) and various types of clustering technologies.

3.1. Storage Virtualization

Storage virtualization plays an increasingly important role as the amount of data that enterprises need to manage grows. Purchasing additional dedicated storage for individual applications leads to inefficiencies and increased complexity of the storage environment. More efficient and cost-effective storage architectures are essential to maximize utilization and to reduce administration costs. The objective of storage virtualization is decoupling logical data access from the corresponding physical resources, allowing the sharing of storage resources across applications, and providing a unified administration interface [Nieboer, 2002]. Potential drawbacks of storage virtualization are reduction in performance resulting from the additional virtualization layer, and lack of direct control over individual storage resources [King, 2004]. In general, storage virtualization solutions require an intermediary system which controls the access from the host network to the storage network. Virtualized storage results in more effective data sharing and reduction in data exchange traffic between clients and servers. This can lead to improved efficiency of server and network resources, better scalability and performance, and potential reduction in costs.

3.2. Server Virtualization (Clustering)

Various types of clustering technologies have been used to improve scalability and availability in mission critical applications, e.g. database management, since 1980s. Computing clusters typically consist of server nodes, high-speed interconnect, and in some situations, a shared disk system. Cluster nodes have their own memory, processors and operating system instances; this allows independent operation of individual nodes, and at the same time creates a large virtual computing resource. Each node can have multiple processor units, similar to SMP (Symmetric Multiprocessor) configuration. In the event of node failure the system automatically initiates fail over operation, assigning the identity of the failed node to a functioning node and resuming the operation. Applications are not aware of the underlying complexity and see the cluster as a single virtual system. Cluster technologies have evolved to provide highly scalable and reliable platforms for enterprise computing, recently entering mainstream computing in the form of blade server technology. Forrester Research estimates blade server shipments to be around 175,000 in 2004, and regards blade servers technology as sufficiently mature to be considered as a suitable replacement for stand-alone servers [Fichera, 2004].

High availability is probably the most important benefit of cluster computing. A cluster remains operational even when individual nodes fail as long as at least one of the nodes remains operational. Up or down scalability is achieved by adding or removing cluster nodes according to processing needs. Adding processing power to a cluster system is cost-effective, and does not require system downtime or expensive memory/CPU upgrades. The workload is shared by multiple nodes of the cluster, improving overall system throughput via parallelism. Administration of clusters is simpler and more efficient than administration of multiple standalone servers, resulting in reduced cost of ownership. Overall benefit derived from cluster technology depends on the type of cluster and level of transparency implemented [Mears, 2004].

Clusters typically operate in two basic modes: active-active, and active-passive configuration. Active-active cluster configuration provides both load balancing and fail over with each node sharing the workload as well as being a fail-over backup for other nodes. This improves both availability and performance taking advantage of parallel processing of cluster nodes. Active-passive cluster configuration does not provide load balancing as one of the nodes remains passive and serves as a backup of the primary server [Moreau, 2005].

3.3 Cluster Architectures

Cluster architectures have evolved as a result of improvements in processor, network interconnect and cluster software technologies. Initially clusters were used to improve performance, availability and reliability of mainframes and supercomputers, but recently cluster technologies have been deployed in smaller systems, including personal computers. Clustering architectures can be classified according to how cluster nodes share resources. *Shared disk clusters* allow all nodes direct access to a shared disk resource. Shared disk is accessed through the system bus to ensure acceptable performance for I/O operations. Moreover, a Distributed Lock Manager (DLM) service is required to control concurrent access to the shared disk. In *shared nothing clusters*, on the other hand, nodes do not share any resource; each node can only access a distinct set of disk resources. When a node fails, control of its storage passes to another active node. Even in situations where the disk resource is shared by multiple nodes, only one of nodes can access it at any given time. The communication between nodes is provided via a network bus or switching fabric architecture. Switching fabric is preferred in the environments that require more bandwidth and less contention.

3.4. DBMS Clustering

Enterprise DBMSs are a popular application for clustering technology and provide one of the early examples of virtualization in enterprise computing. Cluster-aware DBMS is able to utilize underlying processor cluster architecture to parallelize database operations including queries, updates, backup and recovery in a fully transparent manner (i.e. database applications do not have to be modified to take advantage of the underlying cluster architecture). Clustered database systems can be implemented using shared nothing, shared disk or federated database design. Using the *shared nothing* approach data is distributed among cluster nodes, and each node owns its data partition (i.e. has full access to its own data). Other nodes access data through the *owner* node using messaging to perform remote read/write operations. Since data is completely partitioned, changes in the cluster architecture require database reorganization, and in some cases cluster reconfiguration. In stable environments where workload and application requirements do not change frequently, *shared nothing* cluster architecture provides good database performance and availability. In the *shared disk* approach cluster nodes have equal access to all data residing on the shared disk devices. Shared disk clusters can be built using the operating system level abstraction layer that provides transparent access to all database files. As each node can simultaneously read and write any database file, synchronization control mechanism is required to maintain data consistency. This increases the inter-node traffic and may cause performance bottlenecks, in particular in the clusters with large number of nodes. Shared disk clusters provide high levels of availability and scalability, and are highly flexible allowing dynamic workload distribution as new nodes are added and removed. Data is available as long as at least one surviving node remains operational. This is particularly useful during maintenance operations as cluster nodes can be removed and added without taking the database off-line. This approach is favored in real-time applications that require high throughput, reliability and flexibility under the conditions of dynamically changing workload (e.g. e-commerce applications). Oracle RAC and IBM S/390 Parallel Sysplex are examples of commercially available DBMS products that offer shared disk cluster solutions [Mears, 2004].

4 Enterprise Grid Computing

While clustering technologies described in the previous section provide resource virtualization for servers and storage at the operating system and database management level, enterprise applications (e.g. ERP, or CRM) still need to be configured to use specific hardware and software resources. Enterprise Grid Computing attempts to achieve virtualization for enterprise applications. The main idea of enterprise-level Grid Computing is to provide an integrated infrastructure that makes it possible to utilize existing resources to meet workload on demand. Until recently, most Grid Computing applications were scientific, computationally intensive applications [Shimp, 2003]. Such scientific applications aim to solve complex, computationally-intensive problems using resources shared across geographically distributed organizations. Well-known examples of scientific grids include European Commission's EuroGrid that provides resource sharing for Research and

Development projects, USA National Foundation's TeraGrid that supports open scientific research, and NASA's PowerGrid that enables sharing and utilization of high cost resources. While there are some similarities between scientific and enterprise grids, enterprise computing grids have their own specialized requirements. More specifically, they need to be able to support data intensive applications in a reliable and highly efficient manner and provide high levels of virtualization. Effective virtualization requires tight integration between grid components, dynamic provisioning of hardware and software resources, and support for workload balancing [Boucelma, 2002]. Grid computing infrastructure is based on a range of virtualization technologies including clustering (as discussed in section 3 above), storage virtualization technologies such as NAS and SANs, and underlying disk storage technologies such as RAIDs and LVMs (Logical Volume Managers). High-speed Gigabit Ethernet provides fast interconnect to enable reliable data traffic between different nodes of grid. An important aspect of enterprise grid environment is overall management of resources, typically supported by self-adaptive grid management software. The grid management system optimizes distribution of processing among different computer resources and automates system administration continuously monitoring the operations of grid components [Quest, 2002].

4.1 Grid Standards

While hardware and software products for enterprise grid computing are already commercially available, Grid Computing standards are still under development. Global Grid Forum (GGF), the grid standards body, is currently working on the Open Grid Services Architecture (OGSA) specification, which defines services-based approach to grid computing [Hey, 2003]. The latest version of OGSA defines grid architecture with emerging Web services framework and concentrates on knowledge-centric, metadata-driven resource discovery. GGF recently formed a research group for the Semantic Grid project to develop new standards of semantic resource identification and sharing using distributed metadata. The main objective of this project is to enable integration of information from different sources to facilitate efficient grid services [Geldof, 2004], [Goble, 2004]. The OGSA-DAI standard is designed to support data access and integration services in a grid framework based on OGSA Architecture. Using this framework, clients first access the registry to locate a Grid Data Service Factory (GDSF). GDSF returns a *handle* for a suitable Grid Data Service (GDS) that provides access to the identified resource. In addition to supporting database operations (i.e. query, update and load), GDS can communicate with auxiliary services such as the Grid Data Translation Service (GDTS) and Grid Data Transport Depots (GDTD) that provides required data transportation and translation services [Malaika, 2003].

4.2 Data Virtualization Services

As noted earlier, virtualization is essential to the implementation of Enterprise Grid Computing. Virtualization layer hides the complexity and heterogeneity of the underlying hardware and software platforms and in effect creates a *single virtual machine* operating over a *single virtual data source* [Foster, 2003]. To comply with the OGSA specification, grid has to support various levels of virtualization. Data virtualization services hide the complexity, distribution and heterogeneity of grid data sources and present uniform view of the data. Data virtualization involves different types of transparencies that address heterogeneity, location, replication, ownership and distribution of data resources. Combined together, these types of transparencies enable applications to be deployed on the grid without being aware of the underlying data distribution [DAIS, 2003]. Grid data virtualization services are implemented as a layer between grid applications and data sources and function on top of the OGSA services. OGSA virtualization services include data discovery, federated access, workflow coordination and collaboration, and auxiliary services such as schema management, authorization and replication.

5 Conclusions

Major technical challenges will have to be addressed if Utility Computing is to become more than just last year's marketing slogan. Effective resource virtualization in the context of enterprise computing is a necessary prerequisite for attaining the vision of Utility Computing. Virtualization provides a high level of abstraction hiding the complexity, distribution, and heterogeneity of hardware and software resources. Given the complexity and heterogeneity of most enterprise computing environments today, achieving virtualization of computing resources in practice is a major challenge. It can be argued that consolidation leading to major simplification of the enterprise computing environments (as discussed in section 3) should precede any attempts at resource virtualization.

In conclusion, grid infrastructure based on virtualization of IT resources provides the necessary technical framework for implementation of Utility Computing. Leading proponents of Utility Computing, including HP (www.hp.com), Sun Microsystems (www.sun.com), and Oracle (www.oracle.com) are implementing grid infrastructure in their data centers. Sun's N1 initiative, IBM's Autonomic Computing, HP's UDC (Utility Data Centre) all aim at producing large-scale enterprise computing environments based on computing grids and virtualized IT resources. Given this industry momentum it is highly likely that Utility Computing will be the next enterprise computing paradigm.

6 Literature

[Boucelma, 2002] Boucelma, O., et al.: Report on the EDBT'02 Panel on Scientific Data Integration SIGMOD Record, Vol. 31, No. 4, pp. 107 – 112.

[DAIS, 2003] DAIS Working Group.: Services for Data Access and Data Processing on Grids [Online]. Available: http://www.ggf.org/documents/GFD/GFD-I.14.pdf [Accessed 30 Dec 2004].

[Dell, 2004] Dell: Guidebook to the Datacenter of the Future [Online]. Available: http://www.dell.com/downloads/global/products/pedge/en/scale.pdf [Accessed 7 March 2005].

[Fichera, 2004] Fichera, R. Blade Servers Evolving to a Mainstream Technology, Forrester, September 22, 2004, [Online]. Available: http://www.forrester.com/Research/Document/Excerpt/0,7211,35421,00.html [Accessed 27 March 2005].

[Foster, 2003] Foster, I., et al.: The Virtual Data Grid: A New Model and Architecture for Data-Intensive Collaboration [Online]. Available: http://www.griphyn.org/chimera/papers/CIDR.VDG.crc.submitted.pdf [Accessed 10 Jan 2005].

[Geldof, 2004] Geldof, M.: The Semantic Grid: will Semantic Web and Grid go hand in hand? [Online] Available: http://www.semanticgrid.org/documents/Semantic%20Grid%20report%20public.pdf [Accessed 7 March 2005].

[Goble, 2004] Goble, C., De Roure, D.: The Semantic Grid: Myth Busting and Bridge Building [Online]. Available: http://www.semanticgrid.org/docs/ECAISemanticGrid/ECAISemanticGridFinal.pdf [Accessed 7 March 2005].

[Hey, 2003] Hey, T., Fox, G.: Overview of Grid Technology [Online]. Available: http://www.grid2002.org/pgclasssummer03/PGGridPart2jul03.ppt [Accessed 7 March 2005].

[King, 2004] King, E.: The Virtualisation Wars Heat Up [Online]. Available: http://www.windowsitpro.com/Windows/Article/ArticleID/44885/44885.html [Accessed 20 March 2005].

[Kirby, 2004] Kirby, J. P., Klein, E.: Database Consolidation: Reducing Cost and Complexity [Online]. Available: http://www.oracle.com/corporate/ic_library/amr_db.pdf [Accessed 7 March 2005].

[Krantz, Ryan, 2004] Krantz, L., Ryan, J.: Storage Management for Data Center Managers [Online]. Available: http://www.snia.org/education/tutorials/spr2004/san_management/Storage_Management_for_Data_Center_Managers.pdf [Accessed 7 March 2005].

[Malaika, 2003] Malaika, S., et al.: Standards for Databases on the Grid [Online]. Available: http://www.best.tuke.sk/parker/download/oodbs/p92-malaika.pdf [Accessed 7 March 2005].

[Mears, 2004] Mears, J.: Server clusters offer speed, savings [Online]. Available: http://www.nwfusion.com/news/2004/0426cluster.html [Accessed 7 March 2005].

[META, 2004] META Group: First Steps Toward the Data Center of the Future [Online]. Available: http://www.dell.com/downloads/global/corporate/iar/20041101_meta.pdf [Accessed 7 March 2005].

[Moreau, 2005] Moreau, Ken : A Survey of Cluster Technologies [Online]. Available: http://www.hpug.org.uk/ClusterSurvey_-_V2.4.pdf [Accessed 7 March 2005].

[Nieboer, 2002] Nieboer, R.: Virtualization: a strategic tool to beat storage inefficiency - Storage Networking - Industry Overview[Online]. Available: http://www.findarticles.com/p/articles/mi_m0BRZ/is_11_[Kirby, Klein, 2004] /ai_98977134 [Accessed 20 March 2005].

[Quest, 2002] Quest Software : Reducing Downtime During a Consolidation [Online]. Available: http://www.dmreview.com/whitepaper/WID434.pdf [Accessed 7 March 2005].

[Shimp, 2003] Shimp, R. G. : Oracle Grid Computing [Online]. Available: http://www.oracle.com/corporate/ic_library/grid.pdf [Accessed 25 March 2005]

[Subbiah, 2004] Subbiah, V.: Storage Consolidation for Database Environments [Online]. Available: http://www.netapp.com/tech_library/ftp/3311.pdf [Accessed 7 March 2005].

[Yuhanna, 2002] Yuhanna, N.: Oracle Database Consolidation [Online]. Available: http://www.nocoug.org/download/2002-11/db_consolidation.ppt [Accessed 7 March 2005].

[Yuhanna, 2003] Yuhanna, N.: IT Trends 2003: Database Management Systems [Online]. Available: http://www.quest.com/industry_coverage/pdfs/2003db_trends.pdf [Accessed 7 March 2005].