

The reality of a virtual world

Systems run more smoothly and efficiently with virtualization technology. *by Bob Lai*

Virtualization has been the hot topic for the last few years with servers, storage and applications being virtualized. Virtualization promises efficiency, cost savings, increased up-time—and it's even a politically correct “green” effort. Besides gaining an understanding of virtualized environments, the next questions are: “What are the benefits?” and “Can I take advantage of this emerging new trend?”

Virtualization is not itself a goal of astute IT managers, but it is an enabler on which to build automation and a flexible software infrastructure.

What is it?

Virtualization eliminates the need to understand the physical details of the data warehouse technology before it can be used. By eliminating the technical details, the interface and its use are simplified.

For servers, virtualization unlocks the need to run one application using one operating system (OS) on one hardware platform. A virtualized environment allows multiple applications and their corresponding OS hosts to coexist and run on a single hardware platform. By virtualizing the server environment, applications are automatically virtualized—they can now run on any virtualized server. (See figure 1.)

For example, in storage virtualization, the need to know the physical speeds and feeds such as controller interface (SCSI, ATA, Fibre Channel, etc.) and cylinder/

head/sector parameters is completely removed. Normally, storage is accessed through a specific location or a physical address that is a combination of disk hardware parameters. Through virtualization, storage simply becomes a pool of disk bytes (gigabytes and terabytes) from which space is carved out for the applications. Of course, they still have some characteristics that are associated with the storage, such as performance metrics and data protection features. (See figure 2, page 64.)

The benefits

Once all of the underlying technical details are hidden away, virtualization is easy to configure and manage. Also, the associated resources are easy to grow. Since server and storage hardware selection are unlimited, users can mix and match products from multiple vendors and differing models on

a uniform basis without worrying about each system's technology.

For server virtualization, the benefits are, of course, the reduction in the number of hardware servers. This provides savings in equipment costs, maintenance, power and cooling, and floor space. The biggest benefit is the ability to provide all of the necessary processing power to the application as it runs.

It is estimated that most non-virtualized servers run at less than 20% utilization. Virtualizing the server environment increases the data center utilization through a reduction of hardware servers. Without the ability to use the larger processing power of the virtualized server environment, many applications were limited by the power of the server they originally ran on. But now, having fewer servers means less energy is consumed—a definite plus for a green environment.

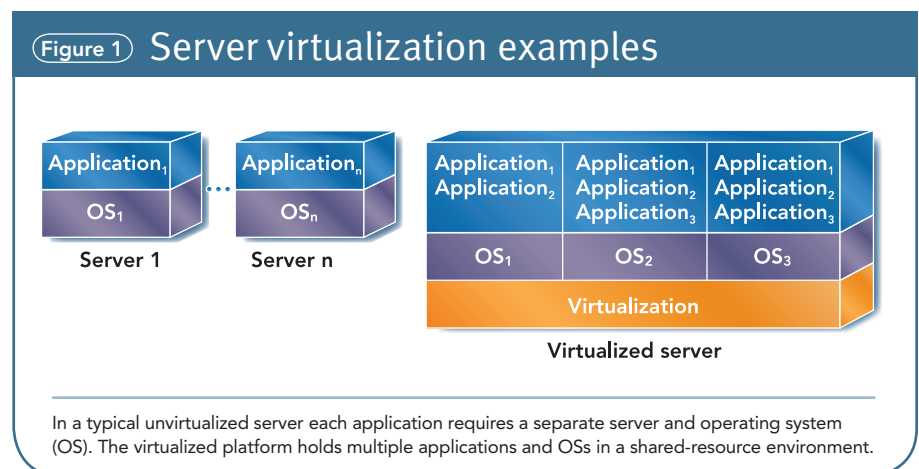
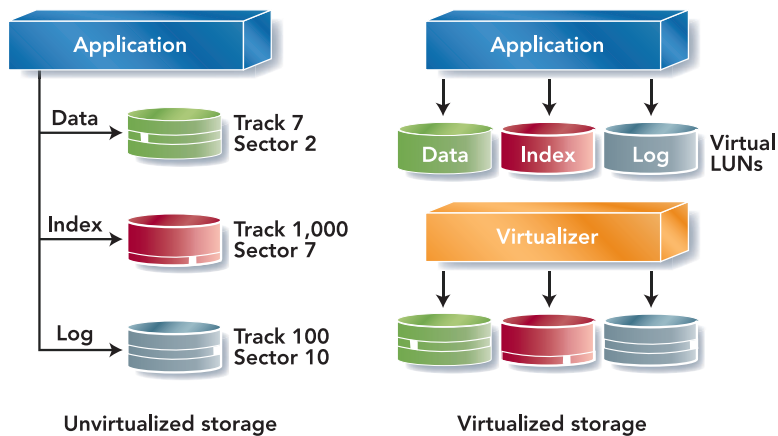


Figure 2 Virtualized and unvirtualized storage units



In unvirtualized storage units, data is located on disk tracks, making it difficult to change. In virtualized storage units, data, indexes and logs on virtual logical units of storage (LUNs) can change as needed.

How to implement

Virtualization for servers, applications and storage is implemented differently, depending on the type that is required.

For server virtualization, a layer of software sits on a server and interacts with the hardware layers. This virtualization layer enables the server to host multiple OS environments; in a non-virtualized environment, each OS requires a separate server. Furthermore, several applications may be running within each OS host environment. In short, what used to run on several distinct servers now runs on one physical server.

Storage virtualization is commonly implemented using a hardware platform specifically designed to control one or more storage systems. From a server perspective, this specialized hardware platform supplies storage capacity for requesting applications. The primary role of the storage virtualizer is to manage the multiple storage systems and provide a seamless single storage image to the servers. Its second role is to hide the management of the underlying technical details.

Storage virtualization solves one of the biggest problems in the data center: the

ever-increasing demand for storage and, in turn, the need for easy administration. Without storage virtualization, each storage increment requires reviewing system specifications to ascertain the type and physical size of the disk, its capacity, the interface required and the chassis location of the next available slot. Knowledge of how to dynamically add a drive to increase

storage capacity is also necessary, as well as the required downtime to back up, expand and restore the system for access. With virtualization, storage expansion is an easy task that can be automatically provisioned without downtime or disruption.

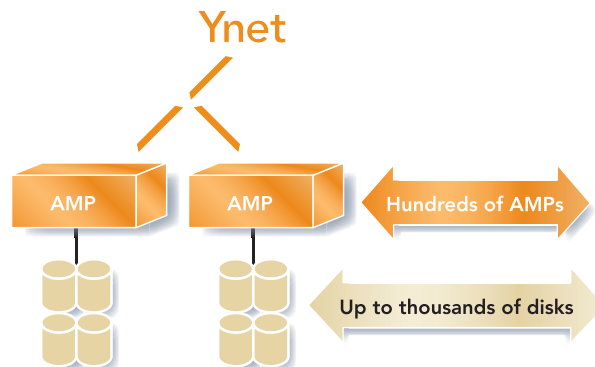
Virtualization and Teradata

Virtualization is now a firmly established theme in the enterprise. Data centers are focused on virtualizing servers, storage, networks and applications. While some segments—especially servers—are easier than others to virtualize, the storage arena is emerging as a critical element of this endeavor. Many efforts to virtualize storage have found it to be a complex process that requires more attention than its counterparts.

However, some applications and solutions have already incorporated storage virtualization technology into their products. One of these is the data warehouse solution from Teradata.

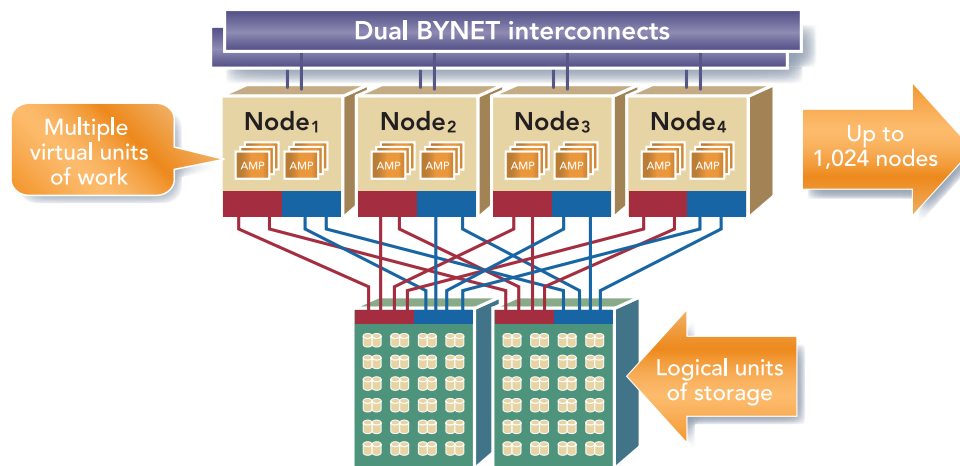
The Teradata solution is an integrated combination of servers, storage and database application software. Since the company's inception, Teradata has continuously tuned and adapted the data warehouse offering for optimal performance

Figure 3 First Teradata architecture, 1985



The Teradata DBC 1012 used a massively parallel processing architecture based on the AMP work unit, which was connected to disk drives and interconnected with the Ynet.

Figure 4 Teradata architecture, since 1995



The Teradata 5500 Server uses a massively parallel processing architecture based on a virtual AMP work unit. These AMPs are abstracted on compute nodes, which are interconnected with the BYNET network, and are assigned local segments of the attached disk array storage for their segments of the database.

and scalability. This has included the use of technologies such as shared memory processor, compute nodes and redundant array of independent disks and, more than 12 years ago, server and storage virtualization.

Teradata uses a massively parallel processing (MPP) architecture that provides a “shared nothing” environment. As shown in figure 3 on page 64, the initial Teradata DBC 1012 system with Teradata Database V1 comprised basic work units, called AMPs, each of which was implemented with a separate processor board and physically connected to a small number of disk drives. Each AMP had to know all about the physical characteristics of where data was placed: sectors, tracks, cylinders and drive. Multiple AMPs were interconnected with the Teradata Ynet network, with each AMP owning its assigned portion of the hashed data tables, to build a scalable data warehouse system.

Over time, evolving processor technologies drove the industry to widely available, powerful servers. To take advantage of these cost-effective computing engines, Teradata transitioned to a virtual implementation of

the AMP work unit in Teradata Database V2. In this environment, multiple virtual AMPs are hosted on a Teradata server element, called a “node.” (See figure 4.) Unlike the physical connection to disks in the original method, these virtual AMPs, often referred to as “VAMPs,” are assigned a portion of the hashed data tables that are stored in the logical storage units on the attached disk arrays.

Virtualization makes the best use of the Teradata compute resources of each node and of the storage resources of the attached disk arrays. A typical Teradata system achieves well over 90% utilization of compute resources. This architecture has been the basis of all scalable Teradata solutions for the past 12 years. With this virtual approach at the node level, the scalability of the Teradata solution is provided by grouping nodes into “cliques” of two to four nodes, each interconnected with the BYNET, to provide overall availability. Multiple cliques are then interconnected to form a Teradata system of up to 1,024 nodes.

The Teradata solution combines server and storage virtualization in its modularly

scalable system. Each node is optimized for the query processing of its data storage component. With more than one node, the computing resources combine globally to handle the data warehouse workload. Similarly, the node’s storage components are globally combined to provide a single view of the data warehouse.

Make full use of your resources

Virtualization is a technology that benefits data centers seeking increased efficiency. Storage virtualization brings an added dimension of ease of use, especially for those environments with rapidly changing needs in capacity and data protection levels. The Teradata solution takes full advantage of virtualization to enable complete use of the data warehouse resources and provide linearly scalable systems to accommodate the requirements of the data warehouse, from entry-sized to the very largest. **T**

Bob Lai is the solutions architect in the Engenio Storage Group at LSI Logic Corporation.