Mainframe Computing Architecture Re-Invented – Open for Business

by Jim Bowman
GE Capital IT Solutions, Calgary, AB

Abstract

The IT industry has re-invented itself several times over the past 20 years, as a result of major changes in computer architecture. We have witnessed the era of mainframe computing, PC's for individuals, midrange clusters for departmental systems, standalone systems for small business, the internet, explosive growth in storage systems and a new growing class of high performance servers. As each shift was introduced to the marketplace, the business community reacted with changing business models. The first wave of dot.coms and e-business is today analyzing the more secure, high performance architectures similar to centralized mainframes. In many ways, we have come full circle.

Sommaire

L'industrie TI a dû se renouveler à plusieurs reprises depuis les 20 dernières années étant donnés les changements majeurs de l'architecture des ordinateurs. On a observé l'évolution des époques de la programmation centrale « mainframe », des ordinateurs PC personnels, des « clusters » mi-niveau pour les systèmes de département, les systèmes autonomes pour la petite entreprise, l'internet, la croissance explosive des systèmes de stockage et la croissance nouvelle des serveurs à haute performance. Au fur à mesure de l'avènement de ces nouvelles technologies, le milieu des affaires a réagi en ajustant son mode de fonctionnement. La première vague des entreprises « dot.coms » et « e-business » analyse maintenant les architectures les plus sécuritaires et performantes similaires au systèmes « mainframe » centralisés. De plusieurs façon, on peut conclure que le cycle d'évolution se referme.

Over the past 20 years, the computer industry has re-invented itself several times. In the late 1970's most business in North America was using a form of mainframe computer architecture which was created by IBM engineers in the 1960's and refined by technology inventors including Honeywell, Burroughs, Digital Equipment, Hitachi Data Systems, and others. These systems used simple operating systems to run a single version of an application to handle mostly accounting applications.

As new technology for chip design was patented, developed for mass manufacturing, it was packaged in small computers for use by small business. This provided an explosive opportunity for more people to become introduced to the benefits of automated accounting, which saved people time and increased the accuracy of information. During the early 1980's, the industry thrived with the creation of new business applications for use in municipal governments, hospitals, classroom education, building construction, and engineering.

The pace of new technology increased again as the Personal Computer model was invented by Apple Computers and IBM. With Microsoft software for Operating Systems, Spreadsheets and Word Processing, individuals could use a computer for daily information. People in all walks of life began creating applications to simplify cooking, writing, homework and any task imaginable.

With the use of 3 different methods of computing: mainframe, distributed and personal systems, many organizations changed their methods of business to achieve more efficiencies, growth and profit. This was also fueled by the business trend to decentralize operations, outsource processing and empower people to run their own piece of the organization.

In the late 1980's, the computer industry created another major advancement with the introduction of UNIX. The concept was to allow application developers a common operating system with which to deliver applications. Using the University of Berkeley kernel, several manufacturer's including SUN, Digital, IBM, HP, and SCO packaged additional support function into this kernel to provide higher levels of reliability, availability and security. During the 1990's, this evolved into common, but unique operating systems that minimized the ability for applications to become portable across different hardware vendors. Many would argue that UNIX has failed to deliver the true heterogeneous model it was intended for. The newest attempt is seen in the LINUX operating system.

The next major breakthrough in computing has been the development of the Internet as a delivery mechanism for computing. As we have seen, the Internet has changed everything, and is evolving rapidly as the main architecture for global communications.

One of the major impacts of Internet usage has been the massive amount of information that is being gathered and created, i.e. DNA and Genome mapping and stored on computers. Industry estimates have stated that the total amount of information in the world will double every 2-3 years. The effect of this is being seen in the business world with introductions of Enterprise Resource Management (ERP), Supply Chain Management (SCM), Customer Relationship Management (CRM), Business Intelligence (BI), E-Commerce and many other major applications that are connecting suppliers and customers together.

The cumulative effect of this explosion of computing demand has caused business and government organizations to begin thinking about centralization of information technology. IT managers are struggling with support, costs are increasing and a general feeling that IT is falling behind, resulting in potential disasters such as the Year 2000 situation and more recently, Sept. 11 in New York.

There are several bright spots in the technology sector which are just now being refined and delivered to the marketplace. These include Copper Chip Technology, Silicon on Insulator (SOI) and Logical Partitioning (LPAR). Each of these technologies offers substantial potential in creating the computing architecture that is required to manage the demand for computing.

IBM has lead the way with the patents and manufacturing development of Copper and SOI. In 1997, fulfilling a dream of several decades, IBM introduced a technology that allows chipmakers to use copper wires, rather than the traditional aluminum interconnects, to link transistors in chips.

Every chip has a base layer of transistors, with layers of wiring stacked above to connect the transistors to each other and, ultimately, to the rest of the computer. The transistors as the first level of a chip are a complex construction of silicon, metal, and impurities precisely located to create the millions of minuscule on-or-off switches that make up the brains of a microprocessor. Aluminum has long been the conductor of choice, but it will soon reach technological and physical limits of existing technology. Pushing electrons through smaller and smaller conduits becomes harder to do – aluminum just isn't fast enough for these new, smaller sizes.

Scientists had seen this problem coming for years and tried to find a way to replace aluminum with the three metals that conduct electricity better: copper, silver, or gold. Of course, if that was simple, it would have been done a long time ago. None of those metals is as easy to work with as aluminum in decreasing amounts. Any new material presents fresh challenges, and reliably filling submicron channels is a bit like filling the holes of a golf course from an airplane.

IBM had to develop a diffusion barrier that could buffer silicon wafers along with the copper. The company has now announced the first commercially viable implementation of Silicon-on-Wafer (SOI) and the ability to apply it in building fully functional microprocessors. SOI refers to the process of implanting oxygen into a silicon wafer to create an insulating layer and using an annealing process until a thin layer of SOI is formed. The transistors are then built on top of this thin layer. SOI technology improves performance over bulk CMOS by 25-35%. It also brings power usage advantages of 1.7 to 3 times, creates higher performance and reliability per processor.

In the past, IBM would spend billions of dollars of R&D, create their own patents and determine what inventions to manufacture based on market dynamics. In 1995, this changed. Today, IBM creates it's own products and also sells the patented technology to it's competitors like, Dell, EMC and Compaq, who use this technology in their products. The result is that the larger computer marketplace is benefiting from these technologies. This allows IT customers to actively plan for systems consolidation of servers and storage that address the processing requirements, reduce IT budgets and improve utilization of scarce human resources.

These technologies were joined by many other advancements in programming, disk storage, software, telecommunications and operations over the past few years. I believe that we are set to achieve a significant improvement in computer performance and will outdistance the mainframe architectures of the past. Mainframes have traditionally been the most powerful, largest capacity and most reliable systems for mission critical computing.

The third major advancement lies in the creation of Logical Partitioning (LPAR) for the new architectures based on copper and SOI.

Why Partition:

The mainframe world has been able to implement partitioning for many years using MVS and Virtual Machine (VM) concepts. These operating systems have not been able to keep up with Windows or open UNIX or the exciting Linux environments. Today, we see a new class of system partitioning available for open systems.

From a marketplace perspective, there is a demand for high end systems to provide greater flexibility, in particular, the ability to subdivide them into smaller partitions that are capable of running a version of the operating system or a specific set of applications.

The main reasons for partitioning a large system are as follows:

- Server Consolidation: Running multiple applications that previously resided on separate physical systems and provide benefits of lower costs, space savings and reduced operation management,
- Production / Test Environments: Partitioning means setting aside a portion of the system resources to be used for testing new versions of applications while the production system continues to run. This avoids buying computers for test applications and provides higher confidence that the test system will migrate to production,
- Increased hardware utilization: Partitioning is a means of achieving better hardware utilization when software does not scale well across large numbers of processors. In such circumstances, running multiple instances of an application on separate smaller partitions can provide better throughput than running a single large instance of an application,
- Application Isolation: Running applications in separate partitions ensure that they cannot interfere with one another in the event of a software failure in one partition. Also, applications are prevented from consuming excessive amounts of resources, which could starve other applications of the resources that they require, and
- Increased flexibility of resource allocation: A workload that has resource requirements that change over a period of time can be managed more easily within a partition that can be altered to meet the varying demands of the workload.

There are several different ways in which server resources may be partitioned. Some vendors distinguish between hard and soft partitions, while others distinguish between physical partitioning, logical partitioning and software resource management.

Physical Partitioning: PPAR - PPAR is used to describe a method of dividing an SMP server into smaller processing units. These units can be considered small servers. It is called PPAR due to the fact that the partitions have to conform to the physical boundaries of the building blocks used to construct the server.

Logical Partitioning: LPAR – LPAR differs from PPAR in the way resources are grouped to form a partition. Instead of grouping using physical building blocks, LPAR adds more flexibility and freedom to select components from the entire pool of available system resources. LPAR works from within a single memory coherence domain, so it can be used within a simple SMP with no single building block structure. All the operating system images run within the same memory map, but are protected from each other by special address access control mechanisms in the hardware, and special firmware added to support the operating system. Several key advantages can be achieved:

- Resource re-allocation for performance,
- Configuration flexibility,
- Full Isolation protection, and
- Fault Isolation.

Resource Management: Many operating systems provide resource management capabilities that can be applied even when the operating system is running within a physical or logical partition. This provides more control over the allocation of computational resources (CPU, memory, I/O) to applications. The allocation of resources can be dictated by different classification rules (users, application, and names). In this way, workloads can be prevented from consuming all the resources. Also, it provides a mechanism to balance the use of system resources optimally. By grouping applications by resource usage behavior, the workloads can be managed together to maximize the utilization of the server.

Now that we have a new class of chip technology available and a good choice of partitioning from the major vendors, it is re-assuring to also see major improvements in the hardware tools for problem management. For years, the mainframe class computers where the only ones which could self diagnose their own problems, providing online reporting back to support staff and even fix themselves. It is exciting to see this same functionality has been brought down to midrange open systems.

Major vendors have introduced concepts of Self-Configuration, Self-Protection, Self-Healing and Self-Optimizing. These sound more like medical terms that computer systems terms, but the concepts are similar. New servers have millions of lines of software code built into the hardware that provides a level of intelligence. Recent functionality in self healing systems include:

- Chipkill ECC memory,
- Cache de-allocation,
- Processor de-allocation,
- Soft memory scrubbing,
- First Failure Data Capture, and
- Redundant bit steering.

When you add all this up, it looks like a new class of mainframe architecture has arrived. Now, the IT staff can seriously examine alternatives for consolidating business applications onto smaller, cheaper, powerful servers, which are easier to manage.

We have witnessed an increased pace of hardware and software announcements being delivered every 12-18 months. It is hard to imagine how much faster the industry will be able to invent new technology for the market, but we will certainly find out.

Stay tuned.

About the author

Jim Bowman began his career with a Marketing / Finance Degree from Northeastern University in Boston. During a decade with IBM, Jim's experience covered sales, marketing, management, training and education, and channel management. Building on a technical background, Jim's focus changed to ERP software with JDA Software in retail and distribution, Labor Management and Shop Floor Data Collection with Kronos and Wireless Integration for SCM. Jim's current role with GE Capital IT Solutions focuses on Infrastructure Consulting for large accounts and e-business implementations. Jim has a fabulous wife, 4 children and enjoys an exciting life in Calgary, Canada.