CLABBY ANALYTICS

Commentary

IBM's New System z10 Mainframe: The Quintessential New Enterprise Data Center

Executive Summary

In late 2002 IBM announced its "on-demand computing vision" — a vision that would change the focus of the entire computing world. In synthesis, on-demand computing called for enterprises to expand the use of information systems — moving beyond the production of inward-looking financial, human resource, manufacturing, sales, and distribution reports, and the operation of mail/messaging systems — into the brave new world of driving *business process flows over service-oriented architecture over consolidated/virtualized underlying information systems*.

In February, 2008, after years of working with customers to build modernized, service oriented, consolidated/virtualized data centers environments, IBM formally articulated how enterprises can build service oriented data centers. Dubbing the new data center model the "New Enterprise Data Center", IBM described how its customers are using technologies such as virtualization, service oriented architecture (SOA), automated provisioning, and business process management software to build data centers capable of supporting tomorrow's interactive, collaborative, Web-based application environments.

> *In addition to announcing its New Enterprise Data Center design, IBM also announced a major upgrade in its System z (mainframe) product line (the z10 EC).*

From Clabby Analytic's perspective (that's me), the new z10 EC represents the ideal system design scale-up, service oriented data center design of the future. Due to an internal high-speed bus architecture, the z10 EC can handle tremendous volumes of messages (service-oriented architecture is "message heavy") — helping to eliminate external networking and drive down networking costs significantly. This type of design reduces much of the need for external switching, for external encryption and intrusion-protection security, and for programming procedure call overhead — as well as reduces associated network latency — and related, associated overhead. Further, reducing external network equipment dependencies reduces energy consumption (myriads of external switches, hubs, and routers — and the energy that they burn — can be eliminated).

> *IBM's System z10, with its tremendous processing power and capacity, high-speed internal bus architecture, SOA (service oriented architecture) tools/infrastructure integration, energy conservation features, and resilience/security/manageability features make it the ideal, quintessential server for supporting service oriented application flow.*

In this *Commentary*, *Clabby Analytics* (that's me), explains why I believe that IBM's System z10 should be every enterprises' "quintessential enterprise choice" when it comes to building a service oriented, New Enterprise Data Center environment.

Background

The enterprise server marketplace is bifurcating — moving away from midrange and low-end distributed tower servers toward high-end, scale-up servers and/or scale-out blades. The primary driver of this trend is a change in the way that applications interact with one-and-other (moving from tightly-coupled, hard-wired to loosely-coupled, message oriented). This change is placing new demands on how information systems process workloads. And, as a result, this change is forcing information technology (IT) executives to reconsider their information systems designs.

To be more precise, for over 40 years IT managers and administrators have been supporting tightly-coupled, hard-wired, oft-times linear, monolithic applications. These applications have usually been written to run on specific operating environments using proprietary interfaces for program-to-program communications and have employed expensive connectors and adapters for data sharing. In these environments, system and application interoperability has traditionally been a major problem. And, applications in these environments are known to be notoriously inflexible — changes anywhere in an application chain can cripple an application workflow. Needless to say, these applications are hard to manage…

Over the past decade, the computing industry has shifted to object oriented development. And industry standards for program-to-program communications (Web services) and data sharing (extensible mark-up language — XML) have evolved. Application development has shifted from building tightly-coupled applications to building loosely-coupled applications that can now request services of one-and-other using Web services and XML standards. Application interoperability obstacles have been mitigated — and, accordingly, a whole new generation of mix-and-match applications has been enabled (as examples, check out how Web 2.0 applications and mash-ups are structured).

This brave new world of "service oriented" applications places different requirements on data centers than the tightly-coupled applications of the past. Service oriented applications are message intensive — so underlying systems/storage and associated networks need to be able to process tens, hundreds, thousands, and millions of service requests in rapid fashion with minimal latency delay. And this is where the sheer brilliance of mainframe and blade system designs comes in…

> *Both mainframes and blades have common design points. Both use high-speed internal bus architectures to message between processors — eliminating unnecessary programming overhead, encryption, and associated network latency delays when trying to communicate to external systems. Both have been tuned and optimized for virtualization (resource pooling) — enabling IT managers to get more processing power out of existing system investments. And both make excellent platforms for consolidation — enabling IT managers to eliminate power-hungry, underutilized, towered application servers.*

The Problems with Current Data Center Designs

Throughout the 1990s and early 2000s, most data centers were designed around distributed computing architecture. IT managers were encouraged to purchase tower and/or rack systems; deploy those systems as distinct application/database/mail servers; underprovision those servers (in order to allow headroom for peak workloads); and then link all of these distributed servers together using 100 MB Ethernet or high-speed Fibre Channels.

The problems with these designs are many. Distributed computing designs:

- *Are hard to manage*. In most geographies throughout the world systems management is the number one cost in operating a data center — often representing 50% of data center operations cost. And there is not a single data center operator in the world who can prove that managing geographically-and/or physically-distributed servers is easier than managing a single server environment. For enterprises looking to reduce the cost of computing, proliferating distributed computing designs is not the answer — consolidating distrib-uted computing designs into more easily managed scale-up environments is.
- *Require a tremendous amount of network hardware and bandwidth.* As stated previously, service oriented applications constantly request and/or provide services. These requests, and their fulfillment, are sent in the form of messages between applications and databases over a network. It makes a lot more sense, when possible, to send these messages across an internal, high-speed bus.
- *Are hard to secure*. Distributed computing designs necessitate multiple access points because they rely on networking tens, hundreds, or thousands of devices together. Scale-up architecture reduces the number of access/intrusion points in a given computing environment — and therefore makes securing that environment significantly easier while reducing security exposure and risk.
- *Create unnecessary programming overhead*. Making a application procedure call between two local processors is far simpler than making a remote procedure call (RPC). Application writers who develop applications for centralized environment actually have to write and maintain less code for centralized, local environments than they have to write and maintain for distributed computing environments.
- *Necessitate redundant purchasing.* In distributed computing designs, IT de-signers usually design systems to fail-over to other separate, discrete systems should a failure occur (meaning additional systems need to be purchased). Scale-up designs, in contrast, fail-over to virtual machines within the same physical systems, obviating the need for redundant system purchases.
- *Waste computing capacity through underprovisioning* (underprovisioning means loading a server to only a fraction of its computing capacity in order to leave room for a given server to deal with increased workload during peak periods). The industry practice for provisioning distributed tower servers is to load these servers somewhere in the 5%-20% range. This leaves at least 80% (but sometimes 95%) of a given server's processing capacity unused most of the time! Underprovisioning computing wastes capacity and energy — and forces enterprises to acquire more computers that do less work than consolidated scale-up servers (mainframes) or consolidated scale-out servers (blades).
- *Waste energy like there's no tomorrow*. Every towered or rack server has its own power supply (power supplies are notoriously inefficient in terms of energy conversion). And every tower/rack has independent fans that are circulating regardless of how hard a given server is working. Further, every distributed tower uses a NIC (network interface card) to communicate externally — and NICs use power. Then add the cost to acquire, deploy, manage and power external hubs, routers, and switches. Combine the power being used by hundreds or thousands of NICs with the power being used by the large number of network devices and it soon becomes obvious that eliminating as many of

these devices as possible can result in significant power (and associated cost) savings.

- *Take up too damn much room.* A mainframe can fit into a typical closet, yet can do the work of several hundred distributed towers (server farms can take up enough real estate to fill an entire house).

Top Ten Reasons Why Mainframes Are the Best Choice for New Data Center Designs

There are ten strong reasons why mainframe architecture is the quintessential system design for next-generation, service oriented, new data center architectures. They are:

1. *Internal network backplane/in-memory processing* — when deploying a service oriented architecture, one of biggest concerns for systems architects is message traffic overhead, followed by the transfer of large files. IBM's System z has a massive internal systems bus capable of supporting large volumes of high-speed network message traffic, as well as capable of transferring large volumes of data between service oriented applications.

   Also, note that the new IBM System z10 can transfer data between processor books (banks of processors contained in a "book" module) at a rate of 6.0 GIGABYTES per second! (This is 2.2x times faster than System z9 — and 15X faster than data transfer in a typical distributed computing environment). Knock-off the overhead associated with issuing remote procedure calls (because procedure calls are issued locally); and knock-off the overhead associated with encryption because local sharing doesn't require network-level encryption) and messaging/ data transfer takes place even more quickly.

   THERE ARE HUGE PERFORMANCE AND SECURITY ADVANTAGES IN HANDLING PROGRAM-TO-PROGRAM COMMUNICATIONS AND DATA TRANSFERS INSIDE A HIGHLY-SECURE, CENTRALIZED COMPUTING ENVIRONMENT.

2. *SOA infrastructure integration* — IBM SOA infrastructure (Web services, XML, automated provisioning, development tools, etc.) is closely integrated with System z operating environments — and optimized for performance on System z architecture. These tools and software products can be found in IBM's Rational, Tivoli, WebSphere, and DB2 product lines.

   Contrast this with SOA infrastructure from other vendors. If you buy from HP or Sun, you most likely will buy a mix of SOA infrastructure and development environments from companies such as Oracle, BEA, and Symantec (VERITAS). There are costs avoidance advantages in buying an integrated stack from a single vendor — especially advantages related to systems, application, and database integration costs.

3. *Advanced virtualization* — Virtualization is the ability to pool and exploit unused physical/logical computing resources. IBM's z/VM operating environment is recognized by users, competitors, and analysts alike as "the gold standard" for virtualization. With z/VM and the z10 "shared-everything" architecture, IT managers and administrators can manage the virtualization of processors, memory and input/output (I/O) better than any other commercial system on the market today — bar none.

   It should also be noted that IBM's z/VM has the ability to run very reliably at 90% or higher system utilization rates (few other virtualization schemes are this power/reliable). Capacity can be ratcheted-up or -down by turning on or off additional processors. And hardware and memory can be easily shared — making it possible to run multiple systems images that can share resources without getting in each other's way.

4. *Energy consumption characteristics* — Start by considering this: in a distributed system environment, every system has its own power supply (often redundant power supplies required for failover). Power supplies are notoriously inefficient — usually losing 25% of the power they receive when converting between AC and DC. Multiply all of the power supplies by all of the underprovisioned servers in a given distributed computing environment — and most IT managers will find that distributed systems are doing less work than can be done on a mainframe — but these systems constantly waste 25% of the energy they receive. System z power supplies are more efficient than most distributed system power supplies — and because IT managers don't underprovision System z's, much more work gets done using less power.

   Next, add in saving attained through consolidation and virtualization. Many accounts are now cutting their server energy bills in half by consolidating and virtualizing server workloads.

   Finally, factor in costs related to network NICs, switches, routers, and hubs. Kiss those devices goodbye, and then tally these savings as part of a System z vs. distributed system energy consumption.

The composite picture is that mainframes are more energy efficient than distributed systems architectures because they: a) don't have to push large amounts of traffic over external networks; b) have greater utilization rates (meaning they don't need to burn as much power as distributed, underutilized server farms); and, c) don't need a myriad of external network devices to support traffic flow.

5. *Manageability* — as enterprises shift from tightly-coupled applications to loosely-coupled service oriented architectures, service interactions — as well as underlying systems/storage/networking functions — will need to be monitored and controlled. IBM offers a bevy of sophisticated business process management, intelligent orchestration (and associated libraries), automated policy-based provisioning, workflow management, service-level management, energy monitoring and management, and dozens of other products that are well-integrated and well-suited to manage service-process flows over underlying, automatically provisioned, virtualized systems infrastructure. Many of these are part of IBM's Tivoli product offerings.
6. *Real-time workload handling responsiveness* — The System z has been architected as a scalable, non-disruptive environment that enables capacity increases to take place transparently to handle unexpected workloads. Distributed system environments usually handle this activity manually — driving-up management costs.
7. *Security* — Distributed computing environment often have thousands of access points; a variety of disparate network, computing, and storage systems; and no centralized point of control. IT departments, accordingly, spend a lot of time and budget locking-down security access points, performing cross-system security integration, on security software site licenses to cover hundreds or thousands of distributed nodes, and on higher salaries and related benefits for IT security personnel.

System z offers IT executives a chance to centralize IT security under IBM mainframe control. Integrated mainframe security can significantly reduce security software licensing and integration costs; greatly reduce your human resource-related security management costs; simplify compliance testing; and eliminate the need to purchase extremely expensive external public keys.

IT executives can expect to save hundreds of thousands (if not millions) of dollars in hardware, software, testing and human resource-related costs by using IBMs mainframe security architecture.

8. *Shared everything (and its ramifications on server availability)* — The System z9 and z10 shares everything including an enhanced I/O subsystem, main memory, and virtualized resources. Important resources are not wasted (as they are in distributed systems architectures). Note: sharing these resources makes failover far simpler in mainframe environments than in distributed computing environments where physical systems are usually assigned to pick-up the pieces when a distributed tower computer fails.
9. *Superior systems and application management* — Advanced systems and applications management programs simplify service management, operations, security, and power management..
10. *Real estate/floor space* — Data centers, particularly in the United States and Western Europe, are starting to run out of floor space. For IT buyers concerned about floor space, the System z packs a lot of processing power into a relatively small footprint (as compared to the floor space that dozens of networked SMP or PC servers might occupy if equivalently configured).

Why the System z10 Is THE Quintessential New Enterprise Data Center Design

Take all of the mainframe advantages articulated in the previous section — and add the following IBM's System z10 differentiators:

- Faster processor speed — As the industry's world's first true *Enterprise* Quad Core processor – one processor can do the work of up to 30 x86 cores. (System z10 quad-cores perform at 4.4 GHz at the chip level);
- New technology delivers on average 1.5 times the performance, with up to 2x for CPU intensive jobs than the z9 EC, and 70% more total capacity than the z9 EC;
- The price/performance ratio has been improved by 35% on IBM's specialty engines (zIIP and zAAP engines);
- Vision to expand System z capabilities with Cell BE ™ technology;
- The ability to integrated new solutions through accelerator technologies enable new classes of applications on System z10.

Further, IBM has recently announced several new products and solutions that make System z10 even more attractive, including:

- z/OS Network Security Services (NSS) - Centralized security services to attached appliance clients;
- New IMS Connect support;
- New DB2 pureXML support;
- Operational Business Intelligence on System z10;

- SAP Business Intelligence Accelerator on System z10;
- Consolidation & Virtualization Server Optimization and Integration Services for System z10;
- SOA Core Banking Renovation & Payments Framework for System z10; and,
- Encryption Authentication - Digital Certificate Authority on System z10.

Summary Observations

Over the last decade or so, the application programming model has shifted from tightly-coupled to loosely-coupled. This loosely-coupled model enables applications to readily interact with each other by requesting and fulfilling services. This approach, combined with industry program-to-program and data-sharing standards has helped overcome interoperability issues and enabled a whole new generation of Web-based, highly-flexible applications to come to fruition.

To support this services-oriented application program model, changes are required in computer systems designs. The traditional model of underprovisioning towered servers needs to be thrown out the window and replaced with a model that emphasizes high resource utilization and services management. IBM's latest "service-oriented data center" announcement portrays a roadmap and blueprint for implementing such a service oriented data center.

As for IBM's System z10 — *Clabby Analytics* believes that Systems z10 represents the quintessential model for service oriented computing. System z10 can be highly optimized (with virtualization to the extreme); IBM's service oriented architecture offerings have been tightly integrated with System z10; and System z10 has several technical advantages over other systems architectures that make it the best, self-contained, centralized service oriented data center in the computing world (bar none). These technical advantages are detailed in the "Top Ten Reasons Why Mainframes Are the Best Choice for New Enterprise Data Center Designs" and the "Why the System z10 Is THE Quintessential New Enterprise Data Center Design" sections herein.

As for parting words of wisdom, consider this:

> *THERE ARE HUGE ADVANTAGES IN HANDLING PROGRAM-TO-PROGRAM COMMUNICATIONS AND DATA TRANSFERS INSIDE A HIGHLY-SECURE, CENTRALIZED COMPUTING ENVIRONMENT. SYSTEMS ARCHITECTS AND DESIGNERS WHO ARE NOT EXPLORING THIS ALTERNATIVE TO EXTERNAL NETWORKING ARE DOING THEIR ENTERPRISES A MAJOR DISSERVICE.*

Clabby Analytics
http://www.clabbyanalytics.com
Telephone: 001 (207) 846-0498

Clabby Analytics is an independent technology research and analysis organization that specializes in information infrastructure and business process integration/management. Other research and analysis conducted by Clabby Analytics can be found at: www.valleyviewventures.com.