

A Consolidation, Virtualization and Power Study on the IBM System z10

June 10, 2008

*Elisabeth Stahl
Michael Buechele
John P. Rankin*

IBM Systems and Technology Group

Executive Overview

As the cost of power grows significantly and data centers start to need more power than they have available, consolidation and virtualization become important energy management strategies for corporations around the world. IBM has long been a champion of energy efficiency and the new IBM System z10 is an outstanding platform for the integration of power efficiency and performance.

This paper will demonstrate the IBM System z10 as an excellent system for power efficiency and performance, discuss server consolidation and virtualization implementation and strategies, and highlight a power benchmarking study on this platform. The paper will conclude with next steps in power benchmarking for the IBM System z10.

IBM System z10 - Outstanding New Platform for Data Center Efficiency

The new IBM System z10 is designed to increase data center efficiency by significantly improving performance and reducing power, cooling costs, and floor space requirements. The z10 offers unmatched levels of security and automates the management of IT resources to respond to changing business requirements.

The IBM System z10 builds on the strengths of the System z platform, delivering new technologies and virtualization that offer improvements in price performance for key workloads. The z10 further extends System z's leadership by expanding scalability for growth and large-scale consolidation, by increasing availability to reduce risk, by improving flexibility to respond to changing business requirements, and by improving security.

The IBM System z10 uses the new Enterprise quad-core processor at 4.4 GHz, up to 1.5 TB of memory, and new connectivity options that enhance its open characteristics. The z10 delivers, in a single footprint, unprecedented performance and capacity growth while drawing upon the rich heritage of previous z/Architecture servers. The z10 meets the needs of large enterprises having large scale, mission critical transaction and data processing requirements while also delivering the scalability and granularity to meet the needs of medium sized enterprises.

The z10 is the industry's only server providing a complete range of policy-driven functions, including:

- Authorization Management to authenticate and authorize who can access specific business services and associated IT resources.
- Utilization Management to drive maximum use of the system.
- Just-in-Time Capacity to deliver additional processing power and capacity when needed to help businesses better manage risk
- Virtualization Security with the U.S. government's highest level of security, Evaluation Assurance Level 5 (EAL5).

For comparable levels of computation, mainframe systems take up dramatically less space. Mainframe computers are also proving to be efficient in cooling and power consumption. This paper will discuss energy efficiency on the IBM System z10 and highlight a power efficiency and performance study using consolidated and virtualized workloads.

The IBM System z10 and Energy Efficiency

The IBM System z10 EC offers a 15% improvement in performance per KWh over the IBM System z9 EC. In addition, the z10 provides monitoring of energy consumption and tools for building energy consumption trends that can be utilized for in-depth analysis. Many corporations are reaching the limits of available space and power at their data centers. With server virtualization and consolidation capabilities and a green footprint, the System z10 is well suited to address these limitations.

The IBM System z10 has a hybrid cooling system that is designed to lower power consumption. It is an air-cooled system, assisted by refrigeration, provided by a closed-loop liquid cooling subsystem. The entire cooling subsystem has a modular construction. Its components and functions are found throughout the cages of the z10.

Several tools are available to monitor the power consumption and heat dissipation of the IBM System z10. The Power Estimation Tool for the z10 is available via the IBM Resource Link Web site (<http://www.ibm.com/servers/resourcelink>). This tool provides an estimate of the anticipated power consumption of a particular machine model and its associated configuration. A user will input the machine model, memory size, number of I/O cages and quantity of each type of I/O feature card. The tool will output an estimate of the power requirements needed for this system which can be used for planning and installing a system. Actual power consumption of the system can be seen on the z10 System Activity Display (SAD) panel. Both the estimator tool and the power display were used in the study discussed later in this paper.

The IBM Systems Director Active Energy Manager (AEM) is an energy management solution building block that controls the energy cost of an installation. It is a cornerstone of the IBM energy management framework and is developed in cooperation with chip vendors, such as Intel and AMD and consortia such as Green Grid. AEM enables clients to manage actual power consumption and resulting thermal loads that the z10 places on the data center. The AEM product is available as a component of Director Services and communicates over a network connection with the client portion on the Hardware Management Console (HMC) of the z10.

Data currently available for viewing on the HMC includes:

- Ambient temperature
- Exhaust temperature
- Average power usage over a one minute period
- Peak power usage over a one minute period
- Status and configuration information

The IBM Systems Director Active Energy Manager provides the intelligence needed to effectively manage power consumption in the datacenter. AEM allows clients to "meter" actual power usage and trend data for any single physical system or group of systems. Developed by IBM Research, AEM utilizes IBM-developed monitoring circuitry to identify how much actual power is being used and the temperature of the system.

With the virtualization and consolidation capabilities, the physical cooling system and the energy consumption monitoring tools, the IBM System z10 has been extensively designed for energy efficiency.

IBM Consolidation and Virtualization as Energy Management Strategies

The IBM System z10 supports the highest levels of consolidation in the industry. Up to 60 Logical Partitions (LPARs) can be used. Each LPAR can run any of the supported operating systems: z/OS, z/VM, z/VSE, z/TPF, and Linux on System z and can be run at up to 100% sustained utilization levels. The z10 is an ideal platform for consolidating many distributed and low utilized systems.

Virtualization is a key strength of the IBM System z10 server. Virtualization is embedded in the z/Architecture, which supports virtualization through both hardware and software. Virtualization creates the appearance of multiple concurrent servers by sharing the existing hardware. Its major goal is to fully utilize resources, lowering the total amount of resources needed and their cost.

Virtualization requires a hypervisor, control code that manages multiple independent operating system images. In System z the hardware hypervisor is implemented in firmware and is called Processor Resource/Systems Manager (PR/SM).

The PR/SM function, responsible for hardware virtualization of the server, is always active and has been enhanced to provide additional performance benefits. PR/SM technology has received Common Criteria EAL51 security certification. Each logical partition is as secure as an isolated server.

Software virtualization is provided by the z/VM product. z/VM uses the resources of the LPAR it is running in to create functional equivalents of real System z servers, which are known as Virtual Machines (VMs). In addition, z/VM is able to emulate I/O peripherals including printers by using spooling techniques and LAN switches and disks by exploiting memory. z/VM's virtualized z/Architecture servers support all operating systems and other software supported on a logical partition. In fact, a z/VM Virtual Machine is the functional equivalent of a real server. z/VM's extreme virtualization capabilities which have been perfected since its start in 1967 make it possible to virtualize thousands of distributed servers on a single z10 server.

System z's virtualization capabilities present a significant opportunity for enterprises to simplify their IT infrastructures. The mainframe's inherent reliability, security and availability as well as its operational model can now benefit other, up to now distributed, applications.

The IBM System z10 also provides advantages in software licensing, since the pricing model for many distributed software products is linked to the number of processors or processor cores. For example, by consolidating and virtualizing under z/VM and exploiting the specialized Internal Facility for Linux (IFL), organizations may achieve a large reduction in the number of used cores, and therefore a reduction in software expenses.

Other uses of virtualization include:

- Isolating production, test, training, and development environments
- Enabling parallel migration to new system or application levels and providing easy back out capabilities
- Supporting back leveled applications

A z/VM production environment can achieve additional savings by:

- Allowing backup virtual servers to be dormant and use no resources until and if they are required. This may help reduce hardware, software and maintenance costs.
- Pooling resources such as processor, I/O facilities and disk space. Virtual servers are provisioned out of these pools and when their useful life ends the resources are returned to the pools and recycled.

- Offering very fast virtual server provisioning. A complete server can be readied for use in just a few minutes, using resources from the pool and image cloning.
- Eliminating the need to re-certify servers for specific purposes. Environments are certified to the virtual server. This needs to be done only once, even if the server requires scaling up, because the underlying hardware and architecture does not change. Significant reductions in time and manpower can be achieved.

IBM is currently conducting a large consolidation and virtualization project internally, which aims at consolidating approximately 3900 distributed servers into approximately 30 IBM System z9 systems. IBM expects to achieve an 85% reduction in IT Data Center square footage and an 80% reduction in energy utilization associated with consolidated servers. Similar results have been achieved by IBM clients and these reductions have directly translated into significant savings. See http://www.ibm.com/systems/optimizeit/cost_efficiency/energy_efficiency for more details.

IBM System z10 Power Study

The IBM System z10 Power Study highlighted in this paper used throughput data for two workloads from the IBM Large System Performance Reference (LSPR) together with power estimates from the z10 Power Estimation tool.

The workloads used in this study were the ODE-B and WASDB workloads from the IBM Large Systems Performance Reference (LSPR). LSPR provides comprehensive z/Architecture processor capacity data across a wide variety of system control programs and workload environments.

LSPR focuses on processor capacity. To assure that the processor is the primary focus, the processor capacity data reported assumes sufficient external resources so as to prevent any significant external resource constraints. With this approach, the LSPR is designed to represent each processor in its best light; the processor itself is the only limiting factor to doing work. Resulting LSPR capacity data is therefore meaningful for establishing a realistic view of relative capacity between specific processor workload environments that have characteristics similar to those measured.

Each individual LSPR workload is designed to focus on a major type of activity, such as interactive, on-line database, or batch. Each LSPR workload includes a broad mix of activity related to that workload type.

Descriptions of these workloads from the [LSPR website](#) are listed below:

ODE-B - On Demand Environment - Batch

The ODE-B workload reflects the billing process used in the telecommunications industry. This is a multi-step approach which includes the initial processing of Call Detail Records (CDR), the calculation of the telephone fees, and the insertion of the created telephone bills in a database. The CDRs contain the details of the telephone calls such as the source and target numbers along with the time and the duration of the call. The CDRs are stored in flat files within a zFS file system. A feeder application reads the CDRs from the files, converts them into XML format and sends them to a queue. An analyzer application reads the messages from the queue and performs analysis on the data. During the analysis further information is retrieved from the relational database, and the same database is subsequently updated with the newly created telephone bill and new records for each call. The feeder and the analyzer applications are implemented as enterprise java beans (EJB) in IBM WebSphere Application Server for z/OS. Using the concept of multi-servant regions, which is unique to the z/OS implementation of WebSphere Application Server, the threads of the feeder and the analyzer applications are distributed over several java virtual machines (JVM). The WebSphere internal queuing engine is used as the queue for the message transport between the feeder and analyzer.

WASDB - WebSphere Application Server and Data Base

The WASDB workload reflects a new e-business production environment that uses WebSphere applications and a DB2 data base all running in z/OS.

WASDB is a collection of Java classes, Java Servlets, Java Server Pages and Enterprise Java Beans integrated into a single application. It is designed to emulate an online brokerage firm. WASDB was developed using the IBM VisualAge* for Java and WebSphere Studio tools. Each of the components is written to open Web and Java Enterprise APIs, making the WASDB application portable across J2EE-compliant application servers.

The WASDB application allows a user, typically using a web browser, to perform the following actions:

- Register to create a user profile, user ID/password and initial account balance.
- Login to validate an already registered user.
- Browse current stock price for a ticker symbol.
- Purchase shares.
- Sell shares from holdings.
- Browse portfolio.
- Logout to terminate the user's active interval.
- Browse and update user account information.

The LSPR Internal Throughput Rate Ratio (ITRR) for multiple z/OS images was used as throughput estimates for the two workloads. The z10 Power Estimation tool was used to estimate the power required for the hardware configurations needed for the various runs. Both estimates were made for configurations with multiples of four processors through the z10 maximum of sixty-four processors.

See the Appendix of this paper for a description of the additional z10 hardware used to generate the LSPR results.

The z10 Power Estimation tool was used to estimate the power required at each point studied. The Power Estimation Tool is a web based tool that allows the user to estimate the power consumption for specific configurations on a specific IBM System z10 machine configuration. Installed processors, memory, features, and I/O are used as input to the tool, which then estimates power consumption for the specified configuration. Note that the tool does not verify that the specified configuration can be physically built. In addition, the exact power consumption for machines will vary; the objective of the tool is to produce an estimation of the power requirements to aid in planning for a machine installation.

A correlation test was performed using real power metrics from an actual IBM System z10 configuration. The same hardware configuration was also used as input to the tool. The results of the z10 Power Estimation Tool were shown to be comparable, within 50 watts, to the z10 power meter display reading.

Table A contains the ODE-B and WASDB LSPR multi-z/OS image ITRRs for the z10 configurations used in the study.

Processors	ODE-B LSPR	WASDB LSPR
	multi-z/OS image ITRR	multi-z/OS image ITRR
4	5.89	5.54
8	11.23	10.39
12	16.20	14.77
16	20.89	18.82
20	25.41	22.69
24	29.77	26.39
28	33.99	29.94
32	38.14	33.39
36	42.22	36.74
40	46.24	40.00
44	50.20	43.19
48	54.12	46.31
52	57.99	49.38
56	61.81	52.38
60	65.56	55.28
64	69.22	58.07

Table A. Workload Internal Throughput Rate Ratios

Table B contains the power estimates for the two workloads on a z10 with the minimum configuration required to run the workload:

processors	ODE-B estimated	WASDB
	WATTS	estimated Watts
4	5,498	5,498

8	5,916	5,566
12	5,984	5,634
16	9,459	9,212
20	9,771	9,277
24	9,837	9,342
28	13,586	13,091
32	13,650	13,155
36	13,901	13,219
40	13,965	13,284
44	17,358	16,681
48	17,420	16,743
52	17,483	16,806
56	17,546	16,869
60	17,936	17,259
64	17,998	17,322

Table B. Power Ratings for Configurations

Chart A shows the LSPR ITRR / Kilowatt plotted against the number of processors. The graphs show that the basic relationship between workload and power is very similar for both workloads. As the number of processors configured in a system increases from 4 to 8 to 12 processors, the capacity per kilowatt of both workloads improves. The LSPR ITRR / KW dips when going from 12 to 16 since the 12 processor configuration is a single book system and the 16 processor configuration requires two books. The additional processors are then added on the two book system going up through 26 processors, giving substantial increases in capacity without significantly increasing the power required. The change from two to three and three to four books is similar. The best efficiency for both workloads comes at 64 processors.

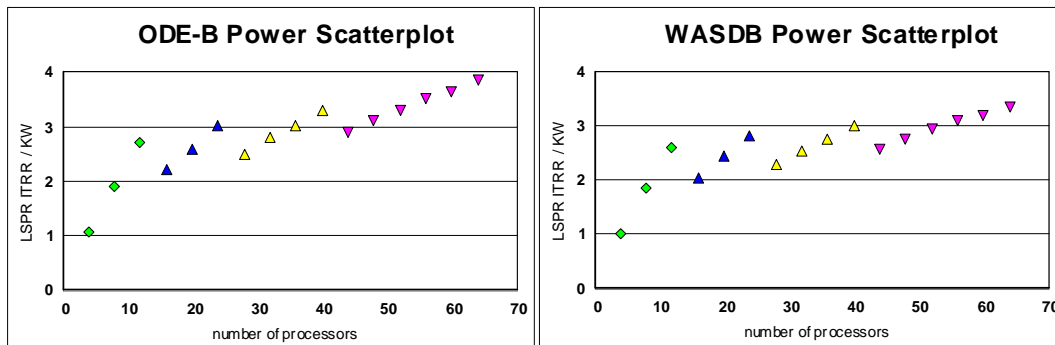


Chart A. LSPR Rating / Kilowatt

As is clear in Table B, very little extra power is required to add more capacity for either type of work so long as the resulting workload fits within a book boundary. More throughput is gained than expended in additional power, and the power efficiency also improves as work is added within a given book footprint. When a small amount of work is added which increases the number of books required, power efficiency decreases slightly. But, especially going from three to four books, adding workload opens up the possibility of additional efficiency at even higher throughput.

For these workloads, this study has demonstrated that consolidating additional work onto existing machines is an effective power performance management technique. This study demonstrates the scalability of the system while power consumption remains relatively fixed; the marginal or incremental power requirements to run extra work are very small. The IBM System z10 mainframe offers a high utilization rate with systems designed to operate at near 100 percent capacity. Based on this analysis, server consolidation and virtualization onto a large server can prove to be an excellent energy management strategy.

Note: IBM does not guarantee that your results will correspond to the ratios provided herein. This information is provided "as is", without warranty, express or implied.

Conclusion

Many organizations around the globe are looking to reduce power consumption and many are facing data center power challenges. With the IBM System z10, clients can successfully integrate reduced power consumption with increased performance.

IBM has led the technology industry in energy-smart innovation for over 40 years and is committed to climate protection. It is IBM's goal to sustain leadership in energy conservation and management by continuing to deliver power-management and cooling technologies. With these technologies, systems use less power, generate less heat and use less energy to cool the system.

Along with advanced technologies, IBM can assist organizations to optimize the utilization of data center and system solutions. IBM Systems and Technology Group Lab Services for System z provides services including power consultations, consolidation assessments, and total cost of ownership offerings. (<http://www.ibm.com/systems/services/labservices>)

This paper has described IBM System z10 energy efficiency strategies including consolidation and virtualization and has highlighted a z10 power study. Consolidating additional work onto existing systems is an effective power performance management technique as demonstrated in this paper. As additional power benchmarking analysis is performed and technologies such as power savings mode are exploited, the IBM System z10 will continue to be viewed as an outstanding platform for power efficiency.

Appendix

Configuration information for WASDB used as input to the z10 Power Estimation Tool:

CPs	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64
Model	E12	E12	E12	E26	E26	E26	E40	E40	E40	E40	E56	W56	E56	E56	E64	E64
CPs + SAPs	7	11	15	22	26	30	37	41	45	49	54	58	62	66	71	75
Voltage Group	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V
I/O cages	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Memory	16	16	16	16	32	32	32	32	64	64	64	64	64	64	64	64
IBT-2-Copper Fan-out	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
IB-MP Daughter Cards	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Ficon Express2 2G LX	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
OSA Express2 1000base-T	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Power Estimate Watts	5498	5566	5634	9212	9277	9342	13091	13155	13219	13284	16681	16743	16806	16869	17259	17322

Configuration information for ODE-B used as input to the z10 Power Estimation Tool:

CPs	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64
Model	E12	E12	E12	E26	E26	E26	E40	E40	E40	E40	E56	W56	E56	E56	E64	E64
CPs + SAPs	7	11	15	22	26	30	37	41	45	49	54	58	62	66	71	75
Voltage Group	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V	208-240V
I/O cages	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Memory	32	64	64	64	96	96	96	96	128	128	128	128	128	128	128	128
IBT-2-Copper Fan-out	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4
IB-MP Daughter Cards	3	5	5	5	7	7	7	7	8	8	8	8	8	8	8	8
Ficon Express2 2G LX	2	4	4	4	6	6	6	6	8	8	8	8	8	8	8	8
OSA Express2 1000base-T	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Power Estimate Watts	5498	5916	5984	9459	9771	9837	13586	13650	13901	13965	17358	17420	17483	17546	17936	17998

References

IBM System z10 Enterprise Class Technical Introduction

<http://www.redbooks.ibm.com/redpieces/abstracts/sg247515.html>

IBM System z10 Enterprise Class Technical Guide

<http://www.redbooks.ibm.com/redpieces/pdfs/sg247516.pdf>

IBM launches New "System z10" Mainframe

<http://www.ibm.com/press/us/en/pressrelease/23592.wss>

Square D PowerLogic Energy Meter

<http://www.powerlogic.com/products.cfm?id=19>

Large Systems Performance Reference for IBM zSeries and S/390

<http://www.ibm.com/jct03004c/servers/eserver/zseries/lSpr/>

Large Systems Performance Reference

<http://www.ibm.com/jct03004c/servers/eserver/zseries/lSpr/pdf/SC28118712.pdf>

S/390 Server Consolidation - A Guide for IT Managers

<http://www.redbooks.ibm.com/redbooks/pdfs/sg245600.pdf>

Power Estimation Tool

<https://www.ibm.com/servers/resourceLink/hom03010.nsf/pages/pet2097v2100>



© IBM Corporation 2008
IBM Corporation
Systems and Technology Group
Route 100
Somers, New York 10589

Produced in the United States of America
June 2008
All Rights Reserved

This document was developed for products and/or services offered in the United States. IBM may not offer the products, features, or services discussed in this document in other countries.

The information may be subject to change without notice. Consult your local IBM business contact for information on the products, features and services available in your area.

All statements regarding IBM future directions and intent are subject to change or withdrawal without notice and represent goals and objectives only.

IBM, the IBM logo, System z, and System z10 are trademarks or registered trademarks of International Business Machines Corporation in the United States or other countries or both. A full list of U.S. trademarks owned by IBM may be found at: <http://www.ibm.com/legal/copytrade.shtml>.

Linux is a trademark of Linus Torvalds in the United States, other countries or both.

Other company, product, and service names may be trademarks or service marks of others.

IBM hardware products are manufactured from new parts, or new and used parts. In some cases, the hardware product may not be new and may have been previously installed. Regardless, our warranty terms apply.

This equipment is subject to FCC rules. It will comply with the appropriate FCC rules before final delivery to the buyer.

Information concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of the non-IBM products should be addressed with those suppliers.

All performance information was determined in a controlled environment. Actual results may vary. Performance information is provided "AS IS" and no warranties or guarantees are expressed or implied by IBM. Buyers should consult other sources of information, including system benchmarks, to evaluate the performance of a system they are considering buying.

When referring to storage capacity, 1TB equals total GB divided by 1000; accessible capacity may be less.

The IBM home page on the Internet can be found at: <http://www.ibm.com>.