Best Practices for Improved z/OS Performance and Lower TCO

By Ivan Gelb
TRADEMARKS

- The following are trade or service marks of the IBM Corporation: CICS, CICS TS, CICSPlex, DB2, IBM, MVS, OS/390, z/OS, Sysplex, Parallel Sysplex. Any omissions are purely unintended.

- © 2019 Gelb I. S. LLC
  URL: www.gelbis.com
  Phone: 732-303-1333
  No part of this material can be reproduced by any means without prior written permission from the author and with proper attribution displayed.
MOTHER OF ALL DISCLAIMERS (MOAD)

- All of the information in this document is tried and true. However, this fact alone cannot guarantee that you can get the same results at your place and with your skills. In fact, some of this advice can be hurtful if it is misused and misunderstood. As with all kinds of analysis, anything you may hear or read can be understood and misunderstood in many ways that may seem contradictory to you. Gelb I. S. LLC, Ivan Gelb and any one found anywhere assume no responsibility for this information’s accuracy, completeness or suitability for any purpose. Anyone attempting to adapt these techniques to their own environments anywhere do so completely at their own risk. 😊 ?
Agenda

- Background
- Best Practices
- IRD
- WLM
- VWLC
- zAAP & zIIP
- z Others…
Background

• Costs in a TCO analysis

A. Environments:
   (1) Production, (2) High availability, (3) Development, (4) Quality Assurance/Testing, (5) Disaster recovery…

B. The Major Costs components:
   (1) People, (2) Hardware, (3) Networks, (4) Storage, (5) Software, (6) Physical facilities, (7) Consumables…
Background : Objective Expert Finding

- Apples-to-Apples studies and projects show that z/OS TCO beats any distributed OS for applications of as little 8 MSUs and greater.
- Larger projects? → Higher percentage savings, not just higher total, $$$ savings.
- Z’s Lower cost is coupled with industry leading attributes of service (RAS + + +)

Data source: Over 100 large scale competitive procurements in private industries and government agencies. Over 300 client sites.
Background : The UOW Cost Pattern

- As Unit of Work (UOW) volume increases:
  - z/OS TCO / UOW decreases
  - Distributed OS TCO / UOW increases
Best Practices / Recommendations

- Consider IBM’s Mainframe Solution Edition program for “recruiting” non-z workloads

- Understand and document business needs baselines, forecast, and seasonal cycles.

- Conduct business focused / aligned Performance Management (PM)

- Conduct business focused / aligned Capacity Planning (CP)
Best Practices / Recommendations - 1

- Effective PM & CP activities will yield required service level quantity and quality for least total cost. Continuously convince your C-management of this!

- Design WLM service policy to mirror business activity – this enables the most effective PM & CP activities. You will always know when you need to do something, and when you may need to pay closer attention.

- Always communicate with management in common business terms: orders, deliveries, effect on total cost.

- Always avoid technical jargon and acronyms when communicating with business people.
Best Practices / Recommendations – 2

- Given less time and more data to analyze, choose your tools and techniques so YOUR effectiveness is improved.
- Practice routine WLM service policy health checks. Ask: “Is it still working as we intended to?”
- Choose tools and techniques that enable analysis of each workload independently and in combination with present and future workloads.
- For capacity planning studies, insure that you isolate workloads not just along business importance but also based on key attributes that affect scalability: IO intensity, virtual storage needs, use of z/OS services, total CPU time in applications code, etc…
BP/R – 3: CICS Regions Consolidation

Production CICS Regions: Consolidate and Make threadsafe

- Recall why each region was created, if you can, and then ask if it is time to put “Humpty Dumpty” back together again.

- Large potential processor savings when fewer regions are run for the same amount of business activity.

- Consider potential benefits of making CICS applications threadsafe for improved exploitation of latest hardware and software improvements. 80/20 rule works very well for this activity!
BP/R – 4: PR/SM Considerations

- If total PR/SM overhead greater than about 1.5%, try to figure out what is causing it and is it worth it.

- Run minimum possible number of LPARS

- Define the minimum required logical processors
BP/R – 5: z/OS Considerations

- Define the minimum required logical processors per z/OS LPAR

- Subject your WLM service policy to regularly scheduled audits/effectiveness reviews

- Perform regularly scheduled performance reviews of all the major subsystems like CICS, DB2, IMS, MQSeries, etc…

- Conduct periodic overnight batch performance reviews – very likely that this work drove your software costs via the R4HA (RIP 2019!)
Intelligent Resource Director (IRD)

Current Expectations:
1. IRD in production use
2. Performing as expected
IRD Opportunity

• Is anyone using it now?
  A: Yes, over 60% of the sites that can.

• Can IRD help us?
  A: Yes – automates shifting the processor and IO capacity among LPARs within a SYSPLEX.

• Can IRD hurt us?
  A: Not if you do your “homework”

• What to do now to use IRD later?
  A: (1) Prepare an effective WLM service policy; (2) Test IRD; (3) Implement it
IRD Overview

- Intelligent Resource Director (IRD) uses: Partition Resource/System Manager (PR/SM), Workload Manager (WLM), I/O Supervisor (IOS), Channel Subsystem, and Parallel Sysplex
- IRD dynamically moves resources to most important work not meeting service level goals
- IRD components:
  1. LPAR CPU management
  2. Dynamic Channel Path Management
  3. Channel Subsystem Priority Queuing
IRD - LPAR CPU management

- LPAR CPU management adjusts LPAR weight and the number of online logical CPUs as required by WLM.
- Works to meet goals of important work as determined from WLM service policy.
- If weight of an LPAR is raised or decreased, its number of online logical CPUs will be changed to maintain a match between logical and physical CPU speed.
- On various LPAR Activity reports, look for fractional number of CPUs as evidence of IRD’s activities.
IRD – Dynamic Channel Path Management

- IRD DCM dynamically moves channel paths through the ESCON Director from one I/O control unit to another
- Must define channel paths as “managed”
- Improves performance by moving bandwidth to where it is needed the most
- Can improve availability – failed channel dynamically replaced
- DCM will consider changes for each LPAR every 10 seconds, but only ONE LPAR will be changed per interval, AND same LPAR will be changed only once / minute
IRD – Channel Subsystem Priority Queuing

- Channel Subsystem Priority Queuing (CSSPQ) sets I/O request priorities
- WLM can change a workloads I/O priority if the work in not meeting goals and I/O activity delays are the reason
- With Enterprise Storage Server (ESS, a.k.a. Shark), WLM can pass a priority to the control unit
- WLM sets priorities (no external user controls) as follows:
  - System work is at highest priority (range is 1 – 16)
  - Highest importance work missing goal
  - Light weight I/O user have higher priority than heavy weights
  - Discretionary work is lowest priority
Workload Manager (WLM)

Current Expectations:

1. WLM service policy delivers the service levels you hoped for

2. You should schedule WLM service policy “checkups” at regular intervals (don’t wait until you must!)
WLM Opportunity

- Can WLM help us?
  A: Yes

- Can WLM hurt us?
  A: Yes, and it is very likely doing it right now!

- What to do now to use WLM better?
  A: (1) Insure that the policy reflects the business priorities; (2) Create effective report classes for low level analysis; (3) Monitor and improve the service policy continuously – especially with velocity goals.

- “Our WLM service policy is perfect because Guru IM-so-Perfect developed it.”
  A: ? ? ? ? OK! Hello! I am your new Guru…The Perceived Perfection passed on while you were looking elsewhere.
WLM Advice

- Recommendation: Create resource groups for any workload you wish to control regardless of processor utilization level.

Example 1: A “SLEEPER” service class can be limited to a maximum of 1 service unit / second rate.

Example 2: A very effective way to keep lower importance work’s processor “Burn” rate lower than it may tend to be without this control.
Variable Workload License (VWLC) Charging Method for Software

Current Expectations:
1. Sub-capacity licensed software with IBM
2. Sub-capacity licensed now with other z-Software Vendors
VWLC Opportunity

- Can we save $ without hurting anything?
  A(nswer): Yes

- How to obtain C-management support for activity?
  A: Demonstrate the estimates of savings possible under your software capacity planning scenarios

- When will ALL software vendors participate in VWLC?
  A: When WE “persuade” them that there is no other way to do business with us.

- What to do now?
  A: (1) Education; (2) Develop software capacity plans; (3) Prepare estimates of savings and costs trends
VWLC Overview - 1

- Variable workload license (VWLC) charging method available in USA since March 2001 for selected IBM software products. Examples: z/OS, COBOL, CICS, DB2, CICS, IMS, MQSeries plus over 25 more.

- Started sub-capacity software licensing trend. Software license capacity can be dramatically lower than installed hardware capacity.

- Concept is still not fully embraced by ALL of the independent software vendors (ISV) on Earth.
VWLC Overview - 2

- Basis for sub-capacity of VWLC products is LPAR utilization

- Monthly charge based on highest (R4HA) rolling 4 hour average by product summed for LPARs w. software present in them

- Product isolation into LPARs for software capacity planning is a potentially cost saving activity

- 5 – 15% monthly software cost savings achieved

- LPAR’s total capacity can be capped via PR/SM to comply with software license agreement
VWLC Overview - 3

• Sub-capacity Reporting Tool (SCRT) assists the determination of LPAR utilization where VWLC products are running.

• For download and details, visit:
  
  ➢ http://www.ibm.com/zseries/swprice/scrt

• Also obtain the WLC Tool from the above URL. It reports the rolling 4 hour average (R4HA) for LPARs.

• Reference:
  Using the Sub-Capacity Reporting Tool, SG24-6522
Current Expectations:

1. Java programs already run on your zSeries
2. Vehicles for Java code: batch, CICS, WebSphere, others?
3. DB2 Version 8 and z/OS 1.6 or higher
4. zIIP and/or zAAP processors installed
zAAP Opportunity

- Can we save $s?
  A: Yes if you have workloads with Java content on z/OS 1.6 running on z890 and z990 processors.

- When to do it?
  A: As soon as you have the required hardware and software levels to support it.

- How to guarantee management support?
  A: Just estimate the initial savings. Later, keep reporting the ongoing savings in hardware and ALL software costs;

- What to do?
  A: Use available tools to estimate potential savings.
zAAP Overview - 1

- zSeries® Application Assist Processor (zAAP), available starting with the IBM eServer zSeries 990 (z990) and zSeries 890 (z890) servers, is a specialized processing unit that provides a z/OS Java™ execution environment.
- Unlike standard CPs, ICFs and IFLs, zAAPs can do nothing on their own.
- Like ICFs and IFLs, zAAPs do not add to software costs processing on the standard CPs.
- zAAPs are designed to operate asynchronously with the general processors to execute Java programming under control of the IBM Java Virtual Machine (JVM).
Prerequisites:

(zAAPs) may be purchased and installed on all z9s, z990 and z890 servers and future follow-on models only.

The operating system must be migrated to:
- z/OS 1.6 (or z/OS.e 1.6), the IBM Solution Developers Kit (SDK) for z/OS, Java 2 Technology Edition, V1.4 with PTF (or later) for APAR PQ86689 must be used and for WebSphere-based Java workloads, WebSphere Version 5.1 or above is required.
zAAP Overview – 3

- zAAPs can offer a number of cost advantages:
  - Low cost of acquisition
  - Significantly lower maintenance costs than the general purpose CPs
  - *No IBM software charges on zAAP capacity*
  - Potential to reduce costs for sub-capacity eligible IBM software because of lower 4 hour rolling average utilization on general purpose CP based LPARs.

- For more information visit: http://www-1.ibm.com/servers/eserver/zseries/zaap/
VWLS and zAAP Illustrated

**BEFORE zAAP:**
- Machine Type: 2084-B16
- Rated @ 647 MSUs
- Sub-Capacity Pricing based on:
  - LPAR A rolling 4hr avg @ 233 MSUs
  - LPAR B rolling 4hr avg @200 MSUs
  - LPAR C rolling 4hr avg @ 267 MSUs
- Rolling 4hr avg of Machine = 547 MSUs
- Average Prime Shift Machine Utilization = 80%

**AFTER zAAP:**
- Machine Type: 2084-B16
- Rated @ 647 MSUs
- Java cycles executed on zAAPs
- New Sub-Capacity Pricing on *reduced* rolling 4hr avg
  - LPAR A rolling 4hr avg @ 213 MSUs
  - LPAR B rolling 4hr avg @ 100 MSUs
  - LPAR C rolling 4hr avg @ 167 MSUs
- New Rolling 4hr avg of Machine = 480 MSUs
- Average Prime Shift Machine Utilization = 50%

Illustration Source: “z890 and z990 zAAP – What it Can Do for You,” By Kathy Walsh, IBM Corp.
zIIP Overview - 1

- **z9 Integrated Information Processor (zIIP)**, available on z9 BC and EC mainframes, benefit DB2 V8 workloads:
  - **ERP or CRM application serving** - For applications, running on z/OS, UNIX, Intel, or Linux on System z that access DB2 for z/OS V8 on a System z9, via DRDA over a TCP/IP connection, DB2 gives z/OS the necessary information to have portions of these SQL requests directed to the zIIP.
  - **Data Warehousing applications** – Requests that utilize DB2 for z/OS V8 for long running parallel queries, including complex star schema parallel queries.
  - **Some DB2 for z/OS V8 utilities** – A portion of DB2 utility functions used to maintain index maintenance structures (LOAD, REORG, and REBUILD INDEX) that typically run during batch, can be redirected to the zIIP.
zIIP Overview - 2

- (zIIPs) may be purchased and installed on z9 BC and EC servers and future follow-on models only.
- Additional requirements:
  - z/OS 1.6 (or z/OS.e 1.6)
  - DB2 Version 8
  - PR/SM enabled
**Z Others**

<table>
<thead>
<tr>
<th>Current Expectations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. z/VM,</td>
</tr>
<tr>
<td>2. IFL,</td>
</tr>
<tr>
<td>3. zBX, and</td>
</tr>
<tr>
<td>4. Specialty IBM database processors</td>
</tr>
</tbody>
</table>
Z Others…

- Consider the industry-unique investment protection IBM offers with the z Series specialty processors.

- Industry leading best products

- z/VM, IFL, zBX, and specialty IBM database processors produce massive $ savings over the competition with:
  - Better performing hardware and software solutions
  - Reduced hardware and software costs due to the efficiencies
In Closing…(alphabetical order!)

- IRD – the final piece required for automation of on demand computing’s performance
- VWLC – opportunity to save 5 – 15% of annual and potentially large one-time software costs
- WLM – the critical success factor for ALL performance management (PM) and capacity planning (CP) activities for lowest TCO
- zAAP & zIIP– opportunity: improve performance of Java and DB2 workloads while you reduce hardware and ALL software costs
- Z Others: z/VM, IFL, zBX, and specialty IBM database processors are industry leading hardware and software alternatives which outperform and under-price distributed OS options.
Need / Want to Know More – 1of2

- 1st: www.ibm.com/servers/eserver/zseries/

- IBM Redbook: Cloud Computing and the Value of zEnterprise. By Elsie Ramos, Kurt Acker, Robert Green, Sebastien Llaurency


- Computer Measurement Group (CMG): www.cmg.org

- SHARE: www.share.org
Need / Want to Know More – 2of2

- **IRD Information:**

- **Washington Systems Center (WSC) documents:**

- **VWLC Information:**

- **WLM Information:**

- Look for IBM Scorpion studies relevant to your needs
Time for…