DYNAMIC CAPPING SOLUTIONS TO REDUCE MLC

– KEY DIFFERENTIATORS –

Perpetual evidence shows that the mainframe continues to play a key role in business transactions and workloads continue to increase. While numbers vary between studies, it is commonly recognized that 70% to 80% or more of business transactions touch the mainframe at some point or another. Another common truth (also illustrated by a number of different studies) is that one of organizations' top priorities or challenges is the reduction of mainframe related costs.

The cost to use the IBM Z mainframe platform is determined by a number of factors, with a significant recurring cost being the Monthly License Charge (MLC). MLC is applied to the use of proprietary IBM software products used in connection with the mainframe, such as the operating system (z/OS), information management system (IMS), customer information control system (CICS), relational database (Db2), etc.

These software products are measured in million service units (MSU), and the MLC is priced based on a peak four-hour rolling MSU average during the month; referred to as "Rolling 4 Hour Average" (R4HA). To help organizations stem the rising costs associated with MLC, capping strategies such as Defined Capacity (DC) and Group Capacity Limits (GCL) are used to restrict or "cap" the number of MSUs that could be used by a single Logical Partition (LPAR) or group of LPARs respectively. However, there are horror stories of those trying to use capping as a way to decrease costs only to cause Service Level Agreements (SLAs) to be missed.

The issue is that DC and GCL lack flexibility. If capped, an LPAR can be constrained meaning that high priority workloads on that LPAR may not have enough capacity to be completed in a timely manner, while lower priority workloads on another LPAR may have excess capacity to complete its tasks. To address this, a few vendors have developed dynamic workload capping solutions to intelligently reallocate MSU from one LPAR to another. In simple terms, this essentially allows a constrained LPAR with high priority workloads to temporarily "borrow" MSU from another LPAR, maintaining SLA compliance without increasing costs.

Dynamic capping solutions help automate the manual effort involved in managing capping limits, while optimizing capacity across LPARs to minimize MLC impact. This allows MSU availability to be allocated to the most important workloads based on utilization needs, workload importance, defined customer policies, and more. With estimates of base Monthly License Charges at 30% of every IT dollar, a reduction of even 2% or 5% can be a significant savings.

Given the significant value these types of solutions can provide, the following briefly looks at two of the key mainframe dynamic capping solutions available today; BMC Intelligent Capping (iCap) for zEnterprise® (Version 2.0) and CA Dynamic Capacity Intelligence (CA DCI) Version 1.5.

Unlike other optimization methods that move workloads to larger capacity LPARs, dynamic capping solutions from BMC and CA continually analyze workload capacity and automatically and dynamically reallocate capacity to where and when it is needed the most. The following is meant to provide a brief introduction to these respective vendor products, highlight key differentiators, and provide some background for areas of consideration.

CA Dynamic Capacity Intelligence (CA DCI) was first released in 2013 as zDynaCap by zIT Consulting, a German based specialist in mainframe pricing and optimization strategies. Today, CA DCI "provides proactive, predictable capacity management for optimizing system resources for prioritized workloads."

First released in 2014, BMC Intelligent Capping for zEnterprise (iCap) "is an automation solution that reduces costs by optimizing resource utilization while keeping business-critical applications running without interruption."

From a conceptual perspective, BMC Intelligent Capping and CA Dynamic Capacity Intelligence both provide solutions that automatically manage changes to defined capacity settings based on workload profiles enabling customers to lower MLC costs. Both also remove the manual effort associated with managing capping limits, using proprietary engines to calculate how MSU is distributed between LPARs.

Inevitably, although implemented differently, there are a number of similarities between these products. For example, both vendor solutions support a phased implementation including: 1) the collection and observation of workload related data; 2) reporting on what the respective solution would have done (i.e., without actually performing any changes; and 3) full functionality of the respective solutions reallocating MSU based on customer settings, current workloads, and more. However, there are a few areas where these products differ.

The following provides some insight into some of these key differentiators.

1 Scheduling of Multiple Policies for Automated Switching

Multiple Policy and Emergency Override Support

Both BMC Intelligent Capping and CA Dynamic Capacity Intelligence solutions allow users to temporarily override active policy settings to help address emergency or unplanned issues (e.g., quickly removing or increasing capping limits to address emergency Disaster Recovery situations). Similarly, both vendor solutions support the creation of multiple policies allowing staff to manually switch the currently active policy.

Automated Scheduling

In addition to a base policy that is effective 24/7, CA Dynamic Capacity Intelligence provides calendar facilities allowing policies to be automatically switched based on dates and times. This allows scheduled policies to temporarily overwrite the base policy to support the use of different policies to address business days and hours, evenings, weekends, holidays, and/or special events (e.g., Black Friday, Cyber Monday, etc.). Policies are automatically switched relieving staff from manual intervention, and allowing organizations to use policies that best fit multiple varied scenarios.

2 Customizing What Workloads are Considered Important and Time Critical

Key Factors

When deciding how to reallocate MSUs to provide optimal performance while maintaining SLAs, the respective proprietary algorithms use a combination of customer specified settings, Workload Manager importance levels, the R4HA, and more. The level of specificity and flexibility available to customers can play a key role in tuning a solution that best fits the customer's needs.

WLM Importance Levels

In z/OS, a Workload Manager (WLM) uses workload classification, service classes, and performance objectives to manage the resources of the system to meet business objectives. Every unit of work that WLM manages is rated by relative importance from 1 (highest) to 5 (lowest), and Discretionary (no importance). In both vendor solution cases, customers are provided with an opportunity to specify what

Workload Management importance levels should be considered low importance allowing for the opportunity to borrow MSU from those workloads.

Two Importance Levels for All LPARs

vs.

In the case of BMC iCap, customers can essentially specify only two levels (i.e., High and Low). Technically, they are only required to specify the WLM importance level that will be designated as low-importance with everything above that considered high-importance. For example, specifying a low-importance of 4 means that importance levels of 4, 5, and D are considered low-importance. This low-importance level is used by the BMC iCap policy for all LPARs and Groups managed by the policy, with the exception where sub-policies might be used (discussed later).

Three Importance Levels for Each LPAR

In contrast, CA DCI currently provides three levels of distinction supporting what it refers to as Business Critical (BC), Time Critical (TC), and Not Time Critical (NTC). This provides more flexibility not only because of the additional level, but because different level specifications can be made for each LPAR managed by the policy. For example, in LPAR-A, importance levels 4, 5, and D may be considered NTC, while in LPAR-B importance levels 4 and 5 may be considered TC and only D is considered NTC. This flexibility allows one LPAR to run more workloads that are considered time critical compared to another that runs more low-importance workloads.

Workload Projections

Both vendor solutions take a number of factors (e.g., priority, importance, R4HA, etc.) into consideration while calculating where and how MSU will be distributed to managed LPARs. Although a highly simplified explanation, BMC iCap considers the current utilization and distributes MSUs altering DC or GCL capacity limits adding a buffer value. By comparison, CA DCI additionally uses the last 15 minutes of data, extrapolating to calculate a 5-minute projection to help better allocate MSU. For example, if there is a downward trend in one LPAR with an upward trend in another, then extra MSU headroom could be assigned to the LPAR with the increasing workload. This helps reduce the need to reassign capacities at a more frequent interval, which can be time-consuming impacting WLM recalculations and HMC access. Even BMC recommends increasing its default recalculation interval as the number of managed LPARs increase.

3 MSU Limits, Ranges and Cost Reduction Targets

MSU Limits

Within their respective policy specifications, both vendor solutions allow customers to create a hard MSU limit that is available across all managed LPARs before workloads need to be capped. These solutions are designed to help reallocate MSU between LPARs to meet SLAs without letting the R4HA increase above these specified MSU limits. The issue however, is what happens when MSU limits are reduced too aggressively?

MSU Flexibility

In addition to an MSU limit (i.e., MAX), CA DCI allows customers to specify a MSU range adding a MINimum MSU value. While the MAX MSU is essentially a cap similar to BMC iCap MSULimit, the CA DCI MIN MSU is essentially a target. This allows organizations to test lower MSU limits, having CA DCI attempt to reallocate MSU to meet the needs of critical workloads across LPARs while attempting to meet this target MIN MSU. However, if this minimum value is too low and SLAs could be in jeopardy, there is still headroom to allow the R4HA to move above that MIN MSU specification. While that means MLC will be higher than hoped, SLAs are not compromised and the total MSU is still maintained below the MAX MSU.

Cost Efficiency	In this scenario, since the R4HA was required to reach above the target (i.e., MIN) MSU to ensure SLAs continued to be met, a new MLC was reached. At this point, it would not be optimal to continue trying to keep the MSU below the MIN MSU specification for the billing period. As such, an AUTOMIN variable is set to the new R4HA, allowing workloads to use that additional headroom. Similarly, an AUTOMAX value helps avoid new peaks during less critical times. Unless this was an unexpected anomaly, the next billing cycle will automatically reset to the MIN/MAX MSU. This mechanism allows organizations to continue attempting to lower MSU while still protecting SLAs from changes that might be overly ambitious.

4 MSU Value and Cost Based Capping

Cost Limits for More Predicable MLC	While the points discussed thus far have been regarding reallocating MSU to reduce costs by maintaining a lower R4HA, depending on the software used within LPARs MLC costs can still be variable. BMC iCap provides a unique facility to specify a value per MSU allowing capping at a particular value. This allows organizations to specify a value that best suits them and cap capacity to remain below more predictable financial limits.
Targeted Use Cases	In a Managed Service Provider scenario, this may be a benefit depending on how contracts are established. This allows costs to be capped to provide a more predictable expense, but still supports MSU transition between LPARs to help provide optimal performance simultaneously.
Caveats	While Cost Limits are a useful mechanism in certain circumstances, care must be taken when assigning any value. For example, the value per MSU can vary greatly depending on LPARs and the various MLC based software running within them. Furthermore, the introduction of Mobile Workload Pricing (MWP) and z Collation Application Pricing (zCAP), and the newly announced Container Based Pricing can exacerbate the issue. In short, with the complexity in calculating a value per MSU it's possible to inadvertently cap an LPAR based on a cost limit even though there are still MSU available.

5 Sub-Groups and Sub-Policies, Their Limitations and Flexibility

Subdivide Managed LPARs	BMC Intelligent Capping allows organizations to create sub-policies within a policy, while CA Dynamic Capacity Intelligence supports the creation of sub-groups within a policy. While somewhat similar in concept (i.e., subdivide managed LPARs into smaller groups each with individually specified MSU limits), there are subtle differences that can have a significant impact on expected goals and outcomes.
BMC iCap Sub-Policies	In general, BMC Intelligent Capping sub-policies can only be one level deep (i.e., no nesting of sub-policies), and require a minimum of two LPARs. While MSU can be reallocated between the LPARs within a sub-policy, they cannot borrow MSU from the main parent policy or neighboring sub-policies; they can only lend MSU to the main parent policy. Depending on the MSU limits set for the main and sub-policies, this could introduce a scenario where LPARs in a sub-policy may be capped even though there are still MSU available in the main parent policy.
CA DCI Sub-Groups	By comparison, CA Dynamic Capacity Intelligence's sub-groups can be nested (i.e., sub-groups within sub-groups), each with their own MIN and MAX MSU ranges. CA

DCI sub-groups can consist of only a single LPAR, and can share MSU across and between sub-groups if necessary; up to their specified MIN/MAX MSU.

Illustrating Sub-Groups

To illustrate an example of the additional flexibility provided by CA Dynamic Capacity Intelligence sub-groups, a mainframe Central Processing Complex (CPC) is divided into five LPARs (LPA1, LPA2, LPA3, LPA4, and LPA5). The main policy (CPCA) includes all five LPARs, enabling the customer to control the usage of the operating system (i.e., z/OS) used by all five. The first sub level includes two sub-groups: GSAS consists of a single LPAR (e.g., LPA5), which uses SAS software (thus controlling the usage of that software product); while GDBA contains the remaining four LPARs (i.e., LPA1, LPA2, LPA3, and LPA4) that each use Db2 software. The next sub level could break GDBA into sub-groups GCICS and GIMS, where sub-group GCICS contains the single LPA3 (allowing control of its CICS software), and GIMS includes LPA1, LPA2, and LPA4 to control their use of IMS software. By defining these groups and sub-groups this way, the use of specific MLC incurring software products (i.e., z/OS, SAS, Db2, CICS, and IMS) can be tightly monitored and controlled.

6 MLC Sub-Capacity Pricing Options Support

Supported Sub-Capacity Pricing Models

Finally, while the feature functionality provided by these dynamic capping solutions helps organizations reduce mainframe related expenses, prospective customers need to make sure the respective solution supports the desired sub-capacity Workload License Charges. For example, Defined Capacity and Groups Capacity Limits are frequently used in conjunction with the Variable Workload License Charge (VWLC) pricing model. However, some organizations are switching to Country Multiplex Pricing and Mobile Workload Pricing to better fit their strategies.

Country Multiplex Pricing and Mobile Workload Pricing Support

CA DCI can coordinate the reallocation of MSU across mainframe CPC boundaries and as such, provides support for Country Multiplex Pricing. Today, BMC Intelligent Capping requires an iCap Master for each CPC and cannot share MSU across BMC iCap environments (i.e., more than one iCap Master per CPC would not be recommended unless for development or testing purposes). As such, its current architecture may limit its ability to support Country Multiplex Pricing.

The concept of dynamically altering Defined Capacity and Group Capacity Limits is not new. In some cases organizations have been manually making changes to help lower mainframe related MLC costs depending on time and dates. Today, there are only a few automated offerings on the market, with BMC Intelligent Capping for zEnterprise and CA Dynamic Capacity Intelligence being the most significantly backed in terms of available mainframe solutions from a single vendor, dedicated research and development, and mainframe related customer base.

While deceptively simple in concept, these products provide automated solutions to intelligently adjust limits to reallocate MSU to best address workload importance maintaining SLAs while providing potentially significant cost savings related to Monthly License Charges. As discussed, while there are conceptually a number of similarities between these products, there are also areas of distinct differentiation that may have an impact on purchase decisions depending on the prospective customer’s requirements. The intent is for this analysis to stimulate conversations around some of the features that are most important to a customers direct needs and situation.

DYNAMIC CAPPING FOR MLC REDUCTION

– A COMPARATIVE ASSESSMENT –

The following provides a look into the functionality provided by the respective vendor solutions (i.e., BMC Intelligent Capping for zEnterprise and CA Dynamic Capacity Intelligence) regardless of cost. Vendor packaging and licensing practices (and customer budget) may significantly impact whether or not the full functionality is available. For example, one vendor may offer features included with the core product that are otherwise licensed options in another vendor's solution.

○ No Support ◔ ◑ ◕ ● Excellent Support

Core Evaluation Factors		
Feature	**BMC**	**CA**
Ease of Installation	◑	●
Architecture	◕	●
User Interface	◕	◕
Phased Deployment	●	●
Supported Pricing Models (Sub Capacity Support)	◑	◕
Policy Definition / Multiple Policy Support	◑	◕
Scheduled / Quick Policy Switching	◔	●
Automated MSU Reallocation (incl. Across CPCs)	◑	●
Predictions / Load Projections for Next Time Period	◔	◕
Defining MSU Limits or Ranges	◕	●
Defining Cost Limits	◑	○
Workload Priority / Importance Levels	◑	◕
Disaster Recovery / Temporary Override	◕	◕

Core Evaluation Factors		
Feature	**BMC**	**CA**
"What if" Simulation Support [*]	◕	●
Data Collection	◕	◕
Application Tuning [†]	◔	◕
Real-time Reporting and Dashboards [‡]	◔	◕
Visualizations[§]	◕	◕
Alerting and Notifications	◕	◕
Audit Logs	◕	◕
Breadth of Portfolio	◕	◕

While this assessment was commissioned by CA Technologies Inc., Zibis Group does not endorse either vendors' solution, rather profiling them in this instance to illustrate several areas of consideration, specific to the use of modern technology and techniques, and towards the adoption of a mainframe cost optimization strategy.

---

[*] BMC requires the addition of BMC Cost Analyzer for z Enterprise (BMC CAzE).
[†] BMC would require third party Compuware Strobe while CA can provide its CA Mainframe Application Tuner (CA MAT) for CA.
[‡] BMC requires the addition of BMC Cost Analyzer for z Enterprise (BMC CAzE).
[§] BMC requires the addition of BMC Cost Analyzer for z Enterprise (BMC CAzE).

ABOUT ZIBIS GROUP

Zibis Group Inc. ("Zibis Group") is a leading source for market research and advisory services on the Information and Communications Technology (ICT) industry and related sectors. We assist vendors, government institutions, academia, non-profits and other key industry stakeholders by providing them with the tools they need to make informed decisions.

Zibis Group is a niche play in the marketplace for customized ICT sector research. We understand that all companies have distinct interests and unique business challenges. We thrive on rising to the occasion and in seeing our clients succeed. Technology is changing and we're here to help.

For more information regarding Zibis Group, please visit **www.zibisgroup.com**.

ZibisGroup
555 Legget Drive, Suite 304 Tower A
Ottawa, Ontario
Canada K2K 2X3
Tel: +1 613-518-8006
www.zibisgroup.com