zSeries Performance: Performance Considerations When Moving to Fewer Faster CPUs

Abstract:

Current processor design continues to deliver more CPU capacity with the same or fewer number of central processors (CPs). This environment results in some customers running their workloads on configurations with fewer CPs.

While this migration can be very successful with workloads enjoying equal or better performance after the move, there are some planning and tuning issues to consider. This paper will discuss the issues involved with reducing the number of processors and will provide guidelines to ensure a successful migration. The information presented is the result of tuning experience at the IBM Washington Systems Center as well as the results of several tests run in a controlled environment.

CPU Capacity:

There are multiple metrics used to measure the capacity of a processor. MIPS (Millions of Instructions Per Second) is the most common. You need to understand the MIPS capacity of the processor as well as the MIPS capacity of each CP to truly understand the performance characteristics of the processor. As an example, if an IBM 2064-116 is a 2,700 MIPS processor, the entire processor is able to deliver 2,700 MIPS. However, since the processor has 16 CPs, each CP is able to deliver 169 MIPS (2,700/16). Therefore the capacity of the processor can be stated as a 2,700 MIPS processor with a single CP speed of 169 MIPS.

The MIPS delivered by the entire processor is an indication of the amount of work the entire processor can deliver. However, some jobs running on the processor can only use one of the CPs at any given time. The amount of work these jobs can do is limited by the MIPS delivered by a single CP and not the capacity of the entire processor. Using the IBM 2064 mentioned above, a single job may be limited to 169 MIPS even if nothing else is running on the processor. This means the processor could be 6% busy, but the particular job is out of capacity on this processor. This is a result of these jobs having a single TCB architecture. This means regardless of the size of the entire processor, these jobs can never use more than the capacity of a single CP. Additional information on single TCB considerations can be found in WSC Flash 9505.

A faster central processor can have very interesting effects on a single TCB typejob. The faster processor allows these jobs to get access to more MIPS. This may be good, this may be bad. It is good if the single TCB architecture job is a high importance job and you WANT more capacity for this job. This may be bad if the single TCB job is high importance but you DO NOT want more capacity allocated to this job. No tuning changes are needed if you want more capacity for the job. This paper will discuss tuning options if you DO NOT WANT more capacity for the high importance job.

The Effects of LPAR on CP Capacity:

LPAR has been a very successful feature of the IBM S/390 product line. It allows a single physical processor to emulate multiple logical processors sharing the resources of the physical processor. The configuration of the logical processors is controlled using the following LPAR definitions:

1. Amount of processor memory allocated
2. Number of logical central processors allocated
3. Amount of processing capacity guaranteed
4. Defining the physical processors as dedicated or shared

The number of defined logical CPs and the processing weight of each LPAR can have a large impact on the performance of the processor. Considerations for setting these values are discussed below.

LPAR introduced the concept of physical and logical processors. The physical CP is the actual hardware processor in the machine. The logical processor is the processor defined to a partition. Each partition can have any number of logical processors defined up to the number of physical CPs on the processor. The operating system running in the partition does not know about the physical processors, it schedules work on the logical CPs. Work does not actually start to execute until LPAR dispatches a physical CP on a logical CP.

LPAR Weight

The weight of a partition is used to identify the percentage of the processor a partition is guaranteed access to. The LPAR hipervisor enforces this value when there is contention for CPU cycles. When the processor is not busy, and the LPARs are not capped, a partition can use more CPU resource than this guarantee.

The LPAR discussions involving LPAR weights in this white paper only apply to busy machines. The LPAR weight is not a factor on a processor which is not busy. The percentage of the processor a partition is guaranteed is the weight of the partition divided by the total weight of all partitions. As an example, suppose a processor has three partitions defined. The first is called WSC1 and has a weight of 20. The second is called WSC2 and has a weight of 30. The third partition is called WSC3 and has a weight of 50. The sum of the weights is 100. The amount of the processor the partitions are guaranteed is determined in the following fashion:

WSC1	20/100 = 20%
WSC2	30/100 = 30%
WSC3	50/100 = 50%

In order for a partition to use the entire portion of the processor it is guaranteed it must have enough:

- Work
- Logical CPs to schedule the work

Number of Logical Processors Allocated

The number of logical CPs is used to specify the maximum amount of capacity a partition can use. The z/OS operating system will dispatch work across the logical central processors. It can not concurrently dispatch more work to run than the number of logical processors defined in the partition.

If the above mentioned partitions run on an IBM 2064-110 (10 central processor), the ideal number of logical central processors for each partition would be:

WSC1:	2 logical central processors
WSC2:	3 logical central processors
WSC3:	5 logical central processors

This configuration would be optimal for performance although it does not provide much flexibility. WSC1 is guaranteed 20% of a 10 way processor (2 physical CPs of capacity). WSC1 has 2 logical CPs. Therefore, in order for WSC1 to get access to 20% of the 10-way processor, each logical processor must have access to a physical processor 100% of the time.

(2) logical CPs x 100% of the time with access to a physical CP = (2) physical CPs of capacity = 20% of the 10 way processor = LPAR weight guarantee

If the physical CP can deliver 250 MIPS, then each logical CP can deliver 250 MIPS.

What would happen if WSC1 was changed to have 4 logical CPs defined? It is still guaranteed 20% of the 10 way processor (2 physical CPs of capacity), but now the capacity must be shared by 4 logical CPs.

(4) logical CPs x 50% of the time access to a physical CP = (2) physical CPs of capacity = 20% of the 10 way processor = LPAR weight guarantee

*If the physical CP can deliver 250 MIPS, then each logical CP can deliver 125 MIP*S.

What happens if the single TCB job requires more than 125 MIPS? It will run fine with two logical processors defined but could have problems with four logical processors defined.

Tuning Options for Systems Running in Basic Mode:

Moving to a faster uniprocessor will generally provide equal or better workload performance on a processor running in basic mode as long as the total capacity of the processor is the same or more than the processor with more, but slower, CPs.

The throughput of the processor can be affected if the system runs a dominant workload which has a single TCB architecture. If the dominant workload is assigned a high dispatch priority and was limited by the speed of the uniprocessor on the previous processor, it can use a larger percent of the new processor since the new uniprocessor delivers more MIPS. This is exactly what many customers expect/want when they move to a new processor. They are trying to relieve bottlenecks to their high priority workloads. There are no tuning changes needed in this environment. Systems which have this situation have latent demand caused by the speed of the slower CP. When moving to the new processor the utilization of the overall processor may be higher, and the CPU seconds used by the dominant workload will be higher. The increased CPU busy time should be matched by an increase in transaction rate.

Some customers have a requirement to prevent the high priority work from dominating their new processor. They must make sure the lower priority work can get access to the new, faster processor. This environment may require tuning changes to achieve this objective. The tuning options are a function of the WLM mode of the system.

WLM Compat Mode:

Systems running in WLM compat mode still use the IEAIPSxx member of parmlib to specify performance options. The time slice option can be used to limit high priority work from dominating lower priority work. The time slice option will dynamically adjust the dispatch priority of a performance group. The following example shows how time slicing could address this requirement. Further information about the time slice option can be found in the *z/OS MVS Init and Tuning Reference*.

IEAIPSxx without time slice:
CICS will always run at a higher dispatch priority than batch with this IPS.

```
PGN=7,DMN=7,DP=F7       /* CICS Performance Group */
PGN=8,DMN=8,DP=F3      /* Batch Performance Group */
```

IEAIPSxx with a time slice:
CICS will run at a higher dispatch priority than batch 80% of the time with this IPS.

```
TSPTRN=(1,1,1,1,*)    /* Time Slice Pattern */
PGN=7,DMN=7,DP=F2,TSDP=F7,TSGRP=1 /*CICS Performance Group*/
PGN=8,DMN=8,DP=F3  /* Batch Performance Group */
```

WLM Goal Mode

Systems running in WLM goal mode use a service definition to specify performance objectives. One of the advantages of WLM goal mode is the ability for WLM to dynamically adjust the dispatch priorities of the work in different service classes. If the high importance work is over achieving it's performance objective and the lower importance work is missing it's performance

objective, WLM can dynamically adjust the dispatch priorities to provide more CPU resource to the lower importance work.

If more control is desired, a resource group can be specified. A resource group can specify a minimum and maximum amount of service available to a service class. It can be used to limit or guarantee service to a service class. Further information about the resource group can be found in the *z/OS MVS Planning: Workload Management.*

Service definition without a resource group:

```
  Service Class = CICSPROD
        Velocity = 50%
        Importance = 1

 Service Class = BATCH
        Velocity = 30%
        Importance = 3
```

Service definition with a resource group limiting CICS to 1000 service units/second:

```
   Service Class = CICSPROD
        Velocity = 50%
        Importance = 1
        Resource Group = RGROUP1

 Service Class = BATCH
        Velocity = 30%
        Importance = 3

 Resource Group = RGROUP1
        Maximum = 1000
```

Tuning Options for Systems Running in LPAR Mode:

The LPAR weight and the logical CP specification needs to be evaluated whenever z/OS is running in a partition.

Since CPU cycles will be consumed by the LPAR hipervisor to manage the physical processor as well as support the defined partitions, it is critical to estimate the reduced processor capacity of a processor running in LPAR mode. The reduced capacity is often minimal on a processor with a small number of active partitions. Newer processors with fewer, faster CPs may be used to consolidate older processors and a natural offshoot of this may be an increase in the number of logical partitions controlled by the hipervisor. When the number of partitions increases, the IBM LPAR/CE tool should be used to estimate the processor capacity of the physical processor.

LPAR Weight:

The LPAR weight guarantees the share of the processor available to the partition. The weight specified for a partition with fewer, faster CPs must be sufficient to provide enough total capacity

to the partition on the new processor. This value is normally a statically specified value on the LPAR definition panel. However, the WLM IRD CPU Management feature allows the z/OS system to dynamically adjust this value to optimize the performance and capacity of a z/Series processor.

Number of Logical Central Processors:
The number of logical CPs determines the maximum amount of capacity available to a partition. While specifying the maximum number of logical CPs to the partition may initially sound attractive, this can have a negative impact on performance when the processor is very busy.

On a very busy processor, it is better to define the minimum number of logical CPs needed to use the amount of capacity guaranteed by the partition's weight. For example, if a partition is guaranteed 50% of a 10 way processor, at least 5 logical CPs must be defined to be able to access 50% of the physical processor.

It is often desirable to define additional logical CPs for a partition to allow the partition to use additional processor capacity if other partitions are not using their fair share. The challenge is how many additional logical CPs can be added to provide maximum flexibility while minimizing any negative performance impact. Determining how to set this value was the purpose of running the performance measurements discussed later in this paper.

The number of logical CPs associated with a partition is normally a seldom changed value on the LPAR definition panel. However, the WLM IRD Vary CP Management feature allows the z/OS system to dynamically adjust this value to optimize the performance and throughput of a z/Series processor.

Time Sensitive LPAR Performance and Capacity Requirements:
Some installations have varying performance and capacity requirements by shift. First shift may need to support a high performance production partition using a subset of the processor. A second partition is often configured supporting a development workload. The physical processor is very busy and the LPAR hipervisor has to enforce the LPAR weights. In an environment like this case the number of logical CPs on the production LPAR would need to be optimized to the LPAR weight.

Typically second and third shifts have different resource requirements. The resource requirements of the development partition are reduced while the online partition, faced with running the batch window, would like to use any excess capacity available on the processor. The production partition causes the majority of the processor busy, as a result the LPAR hipervisor does not have to enforce the LPAR weights. In this case, the production partition would like to have access to more logical CPs to use all available capacity of the processor.

The following logical CP definition and manipulation can address this environment. Define the number of logical CPs on the production partition to match the number of physical CPs on the processor. This will give the production system access to the entire processor during second and third shifts. At the beginning of first shift use the MVS config command to configure off the

number of logical CPs on the production system, keeping the number of logical CPs optimized to the partitions weight. . This method can be used whenever there is a shift change which would require a change in the logical CP definition, such as holidays, weekends, or periods of low utilization.
This technique is implemented using the operations staff or preferably automation to configure CPs online and offline. The IRD feature of WLM can also be used to have WLM manage this value based on the importance of the work in each of the partitions.

WSC Performance Test:

Customers have expressed reservations about consolidating partitions, running on processors with multiple physical CPs onto a new processor with fewer, faster processors. The consolidation often increases the logical processor to physical processor ratio from the recommended 2:1 range to a 4:1 or greater range.

The WSC tests were intended to better understand this environment and provide guidelines on configuring multiple partitions on a processor with a limited number of very fast CPs.

Hardware Environment:

IBM 9672-RX3	One partition with one dedicated CP
IBM 2064-116	Ten partitions with one CP in the LPAR shared pool
	One partition with fifteen dedicated CPs
IBM 2105 SHARK	DASD
500 3270 Terminals	Simulated by TPNS with a five second user think time

Software Environment:

OS/390	Version 2 Release 10
CICS/TS	Version 1 Release 3
TPNS	Version 3 Release 5

Test Overview:

The transaction rate was controlled using the user think time in TPNS. Initial measurements determined a transaction rate which would drive the 9672-RX3 physical processor to greater than 80% busy A five second user think time was used to accomplished this with a consistent rate of approximately 35 transactions per second.

CICS transactions and TPNS were the only workloads running during the initial phase of each test. Low priority batch work was submitted prior to the end of the test to push each partition to 100% busy.

The PCR (Basic Mode CPU comparison) and LPAR/CE (LPAR overhead) tools were used to estimate the capacity of an IBM 2064-1C1 running in LPAR mode. This single CP was defined to support multiple partitions. The workload in the 9672 partition was replicated into each partition of the 2064.

The partition on the 9672 was using 1 dedicated engine. This engine was 80% busy. When the work was replicated into the ten partitions of the 2064, the single shared CP of the 2064 was also 80% busy.

In order to compare these utilizations the relative MIPS capacities of the processors need to be estimated. For this analysis the 2064-1C1 was chosen as the base processor and the processor capacity set to 250 MIPS. Using this base value, PCR estimated the speed of a single CP on the 9672-RX3 to be 18 MIPS. The LPAR/CE tool estimated the speed of the 2064 single CP to be 201 MIPS when supporting the 10 partitions.

Processor	Mode	MIPS per CP	Comments
2064-1C1	Basic	250	Base processor for comparison
9672-RX3	LPAR	18	Dedicated CP estimated by PCR
2064-116	LPAR	201	1 shared CP for ten partitions

Test Objective:
The objectives for the test were relatively simple and straightforward. They are listed below:

Can multiple images running on multiple slower processors be consolidated onto a single faster processor and maintain the transaction rate and response time?

What impact will 100% busy have on the single, faster processor, with so many partitions to manage?

Test Results:

Test Environment	Physical CP Busy	CICS Trans Rate	CICS Average Trans Response Time (sec.)
9672-RX3 CICS only	84%	35	.101
9672-RX3 CICS plus Batch	100%	34	.097
10 Partitions sharing 1 2064 physical CP CICS only	8.3% per Partition	31 per Partition	.020
10 Partitions sharing 1 2064 physical CP CICS plus Batch	9.9% per Partition	31 per Partition	.026

Conclusions:
The results of these tests indicate it is possible to consolidate multiple partitions or processors which use multiple, slower central processors onto a processor with fewer, faster physical processors.

This does not mean planning is not required to be successful. The configuration of the new processor must consider the following if performance and throughput are to be maintained on the faster processor.

The LPAR partition's weight must guarantee an equivalent or greater amount of CPU capacity on the new processor versus the original processor.
The weight on the test configuration guaranteed 10% or approximately 20 MIPS for each partition. The workload had access to approximately 18 MIPS on the 9672-RX3.

The number of logical CPs defined for the partition on the new processor must provide an equivalent or greater number of MIPS to each logical CP to prevent a reduction in MIPS to workloads having a single TCB architecture.
The test configuration provided approximately 20 MIPS to each logical central processor, while the 9672-RX3 provided approximately 18 MIPS per logical central processor.

It is recommended any processor running in LPAR mode use the appropriate tool(s) to estimate the potential capacity reduction due to the LPAR configuration.
The basic mode capacity numbers are overstated if the processor will run in LPAR mode. This is more of a factor as the number of partitions increase. The IBM LPAR/CE tool in conjunction with the IBM PCR tool provides processor capacity in an LPAR environment. Your IBM account team can run these tools for you. LPAR/CE showed a 20% reduction in 2064 CP capacity used in the test configuration. This reduction is a combination of LPAR busy (*physical) plus elongated TCB time in the partitions. Actual CP capacity reductions are totally configuration dependent.

Special Notices