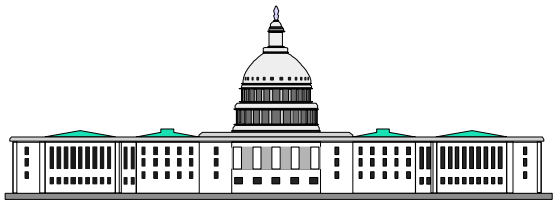
IBM Advanced Technical Support

LPAR Concepts

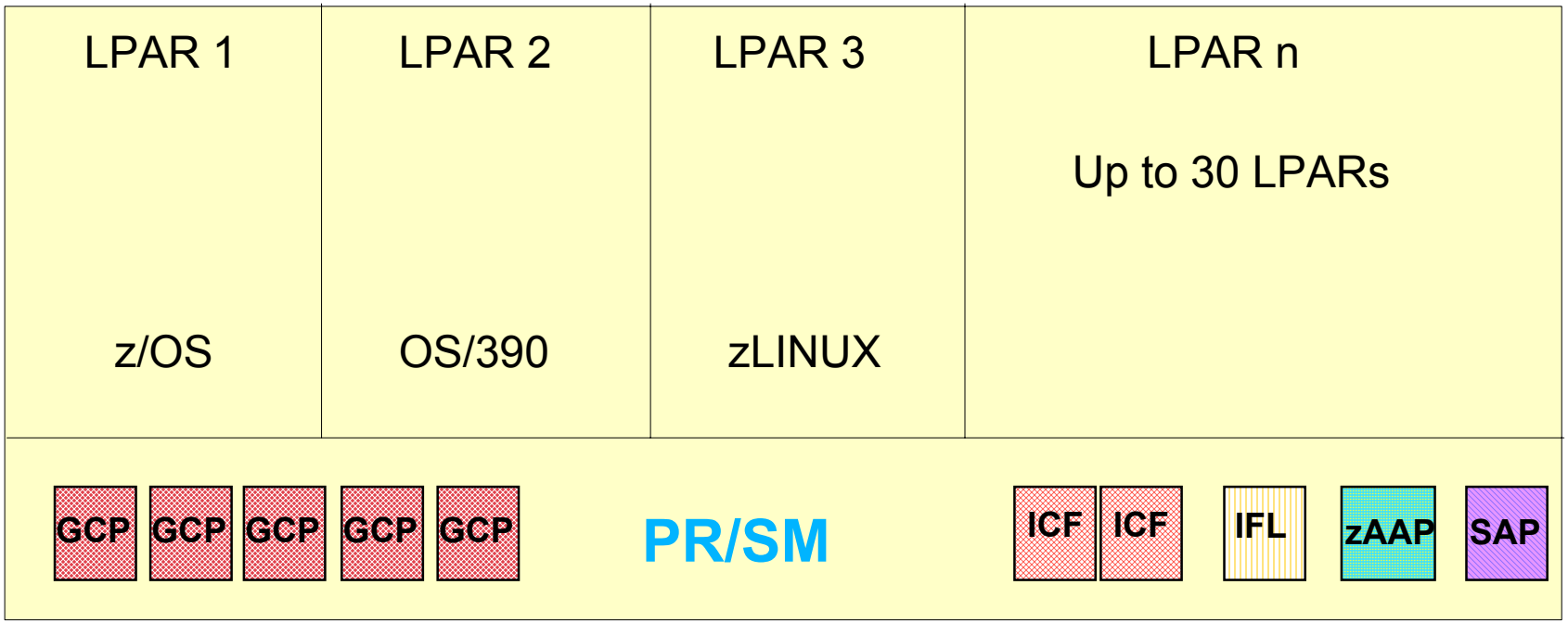
Walt Caprice
Washington Systems Center



Agenda

- Introduction to PR/SM ™ and to LPAR Controls
- "Short" CPs
- Dispatching Work
- Fewer, Faster CPs vs. More, Slower CPs
- Sources of LPAR Overhead
- Miscellaneous LPAR Information
- Capacity Planning Impacts of LPAR

zSeries Virtualization via PR/SM Technology

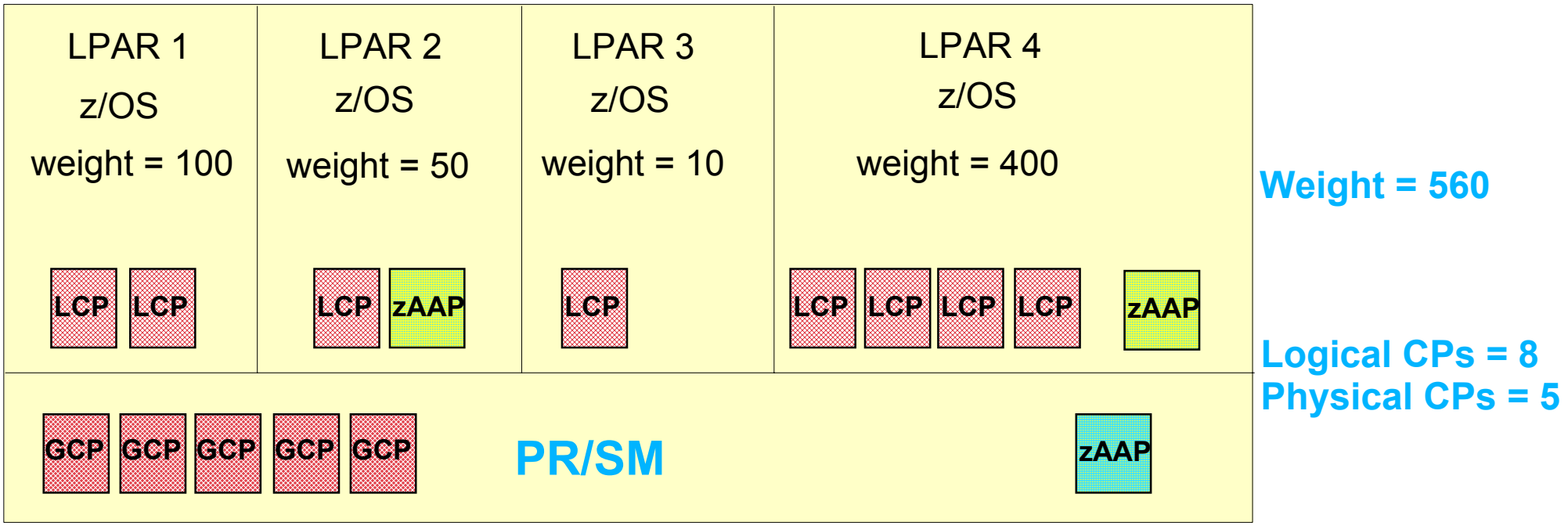


- 1 to 30 LPARs per CEC
- 1 to 32 PUs per CEC (2084-D32)
- Operating Systems don't know they are not running directly on the hardware
- PR/SM ™ is managing the resource allocations based on installation controls
  - PUs can be defined as shared among the LPARs or dedicated to a specific LPAR

GCP	General Purpose CP
ICF	Coupling Facility CP
IFL	Integrated Facility for Linux (VM)
zAAP	zSeries Application Assist Processor
SAP	System Assist Processor (IO)

Partitioning Controls

- Number of partitions, their relative weights, and CP mode (dedicated or shared)
- Number of logical CPs defined to the partitions
- Ratio of logical CPs to physical CPs defined
- Effect of the partitions shared weight and any impact of capping on the partition
- CP usage; either general purpose, traditional CPs or the use of IFL / ICF / zAAP CPs
- Type of system control program (z/OS, z/VM, Linux, etc.), and the workload characteristics in each partition



Important Terms to Understand

- **LPAR weight and per CP share**
- **Effective Dispatch Time**
- **Partition Dispatch Time**
- **Short CPs**

Important Concepts to Understand

- **LPAR weights become important only when the processor is very busy or capped**
- **There are two dispatchers involved in making resource allocations**
  - PR/SM
  - Operating System

RMF Partition Report

```
P A R T I T I O N  D A T A  R E P O R T

MVS PARTITION NAME                    WSC1
NUMBER OF CONFIGURED PARTITIONS          2
NUMBER OF PHYSICAL PROCESSORS            9
                    CP                   9
                    ICF                  0
WAIT COMPLETION                         NO
DISPATCH INTERVAL                    DYNAMIC

--------- PARTITION DATA ---------     PROCESSORS
  NAME     STATUS  WEIGHTS  CAPPING     NUM  TYPE
WSC1          A       800       NO       9  CP
WSC2          A       200       NO       9  CP
```

2084-309 = 325 MIPS/CP

Logical Processor Utilizations

- **Measurement which states the busy of the logical CPs**
  - Independent measure of capacity
  - Can run out of logical CP capacity before the processor is 100% busy

Physical Processor Utilizations

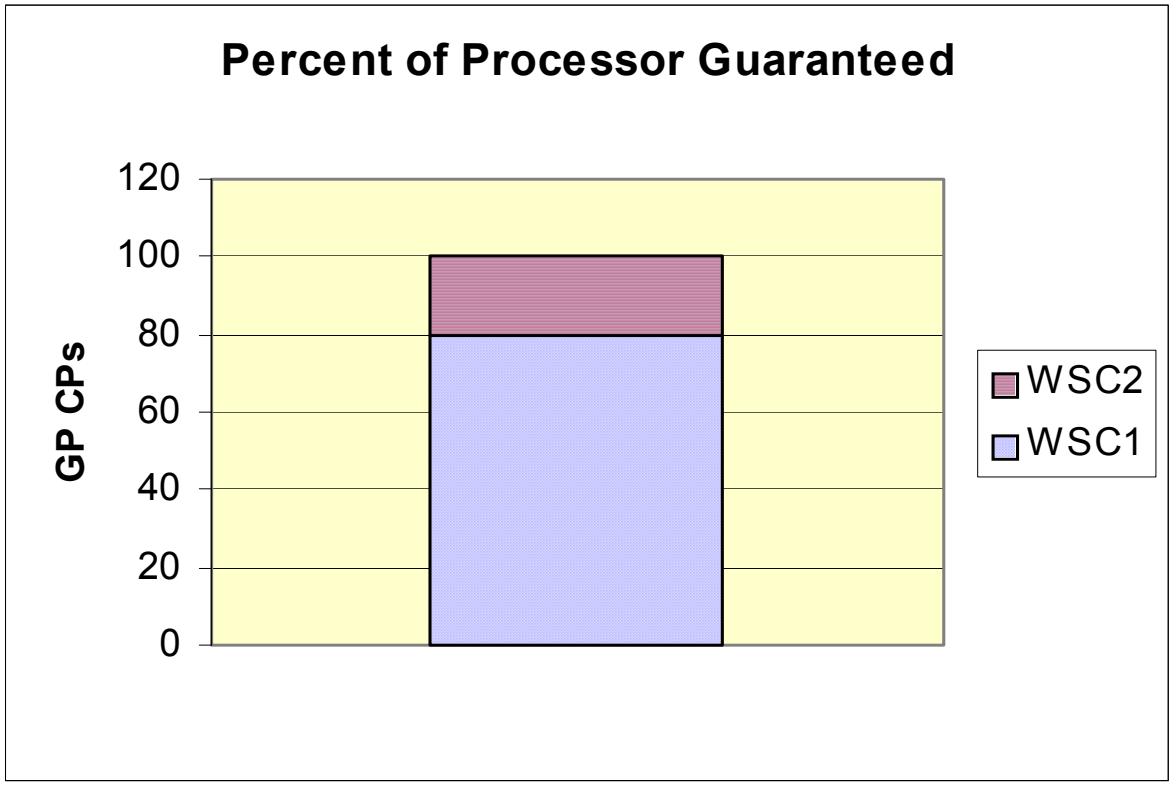
- **Measurement of the partition busy in processor terms**
  - Differs from effective time when the number of logicals defined to the partition does not match the number of general purpose CPs
  - It is this metric which is used in capacity planning exercises

---------- PARTITION DATA ----------									-- AVERAGE PROCESSOR UTILIZATION PERCENTAGES --				
			----MSU----		-CAPPING--		PROCESSOR-		LOGICAL PROCESSORS		--- PHYSICAL PROCESSORS ---		
NAME	S	WGT	DEF	ACT	DEF	WLM%	NUM	TYPE	EFFECTIVE	TOTAL	LPAR MGMT	EFFECTIVE	TOTAL
OSP1	A	100	0	80	NO	0.0	4	CP	19.61	19.62	0.00	4.90	4.91
OSP2	A	100	0	80	NO	0.0	4	CP	19.61	19.62	0.00	4.90	4.90
OSP3	A	100	0	80	NO	0.0	4	CP	19.61	19.62	0.00	4.90	4.91
OSP4	A	120	0	95	NO	0.0	4	CP	94.74	94.75	0.00	23.68	23.69
CF01	A	DED	0	100		0.0	3	CP	99.98	99.98	0.00	18.75	18.75
CF02	A	DED	0	100		0.0	3	CP	99.98	99.98	0.00	18.75	18.75
PHYSICAL											0.02		0.02
											------	------	------
TOTAL											0.03	75.88	75.93

Calculate LPARs Weight (relative value)

$$Share = \frac{LPAR\ Weight}{Sum\ of\ Weights}$$

WSC1 share = 800/1000 = **80%**
WSC2 share = 200/1000 = **20%**

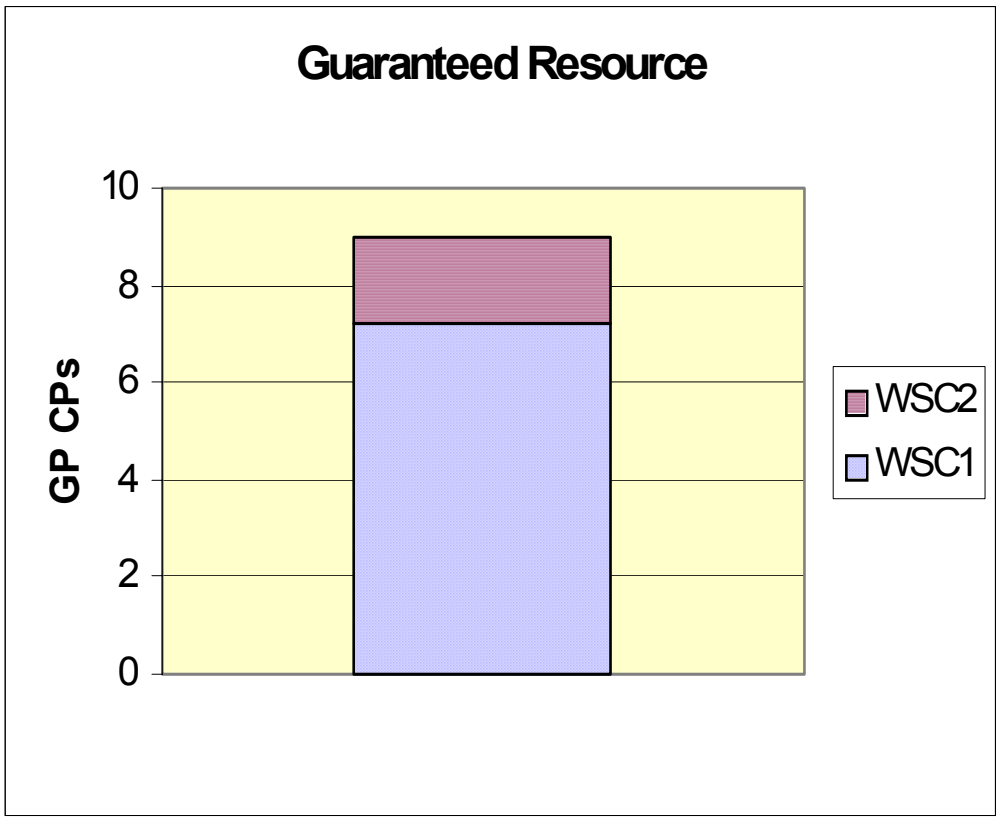


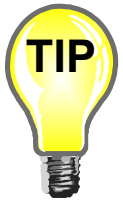
- All active LPARs are used even if an SCP is not IPL'ed
- Only LPARs with shared CPs are used in the calculation

Calculate amount of processor guaranteed to each LPAR

Processor guaranteed =
# of General Purpose (GP) Physical CPs (PCP) * LPAR share

WSC1 capacity = 9 * .80 = 7.2 CPs
WSC2 capacity = 9 * .20 = 1.8 CPs





- The processor guarantee is used to offer protection to one LPAR over other busy LPARs demaning service

Determine Per CP Fair Share Dispatch Time

$$\textit{Partition dispatch time \%} = \frac{\textit{Guaranteed Processor Value}}{\textit{\# LCPs in the partition}} * 100$$

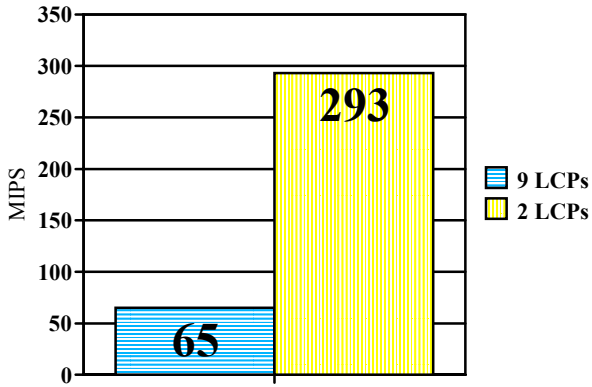
WSC1 = 7.2 / **9** = 80% or .8 * 325 = 260 MIPS

WSC2 = 1.8 / **9** = 20% or .2 * 325 = 65 MIPS

Better Alternative Is:

WSC1 = 7.2 / **8** = 90% or .9 * 325 = 293 MIPS

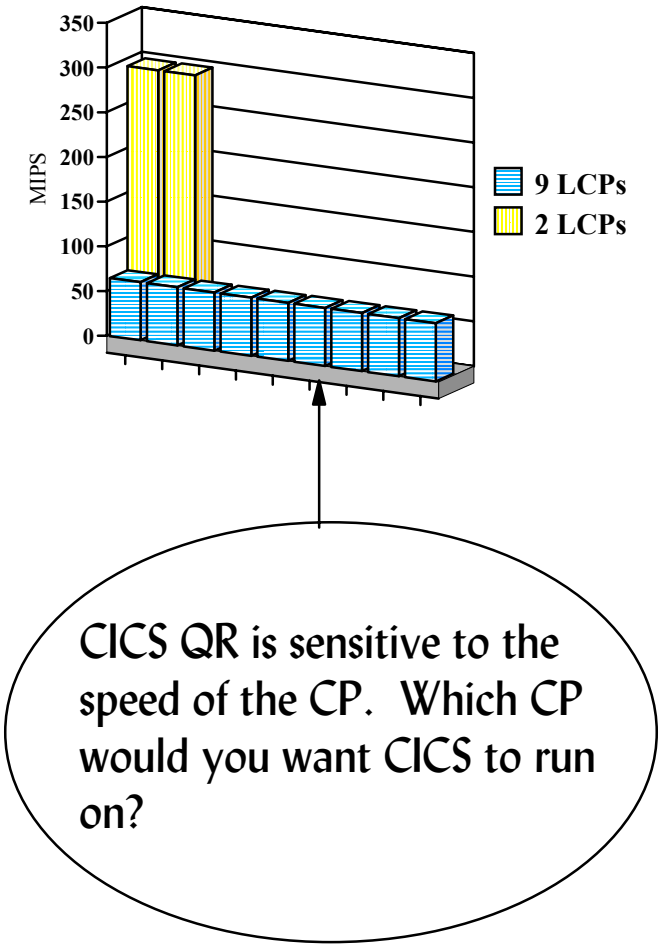
WSC2 = 1.8 / **2** = 90% or .9 * 325 = 293 MIPS



- Biggest per CP Share possible is best when processor is busy

What Are 'Short CPs'?

- **Term created by the WSC performance staff**
  - Performance phenomenon created by LPAR hipervisor enforcing LPAR weights on busy processors or capped partitions
- **LPAR ensures each partition has access to the amount of processor specified by the LPAR weight**
  - This can reduce the MIPS delivered by the logical CPs in the partition
  - Controlled by a combination of LPAR weights and number of Logical CPs
  - Potential Performance Problems
- **In a processor migration "short CPs" are not a problem as long as the partition on the new CEC has access to an equal or greater number of MIPS per CP**
  - Techdocs Item: WP100258 – Performance Considerations when moving to Fewer, Faster CPUs



WSC 'Short CPs

- **Possible Performance Symptoms**
  - Prod CICS can't keep up and transactions are backing up
  - Production system is 'sluggish'
    - **High performance address space may not be getting enough cycles (GRS, XCF, Catalog, etc.)**
  - Test system is not processing XCF requests in a timely fashion and production system is experiencing the performance problem in a Sysplex (*sympathy sickness*)
    - **GRS on production system**
    - **Catalog processing**
- **Due to logical CP losing access to physical CP**
  - z/OS is NOT AWARE the CP is gone
  - High priority task doesn't have a physical assigned, while the low priority task does have a physical CP assigned

Do I Have Short CPs?

- Compare LPAR Busy % versus MVS Busy % on RMF CPU Activity Report
  - If MVS Busy is greater then LPAR Busy 'short CPs' exist
    - May or may not be great enough to cause pain (yet)
- Can calculate the MVS to LPAR busy Ratio
  - MVS Busy / LPAR Busy 71.55 / 31.64 = 2.26
    - Most problems are noticed with a ratio greater than 1.25

```
            z/OS V1R4                SYSTEM ID SYSA              DATE
08/06/2003
                                     RPT VERSION V1R2 RMF        TIME
08.56.59
CPU  2064   MODEL  216
CPU     ONLINE TIME   LPAR BUSY      MVS BUSY      CPU SERIAL  I/O TOTAL
NUMBER  PERCENTAGE    TIME PERC      TIME PERC     NUMBER      INTERRUPT

RATE
0       100.00        31.85          71.76         010B2E      14.09
1       100.00        31.62          71.54         110B2E      20.42
2       100.00        31.57          71.48         210B2E      19.94
3       100.00        31.50          71.42         310B2E      21.84
TOTAL/AVERAGE         31.64          71.55                     76.29
```

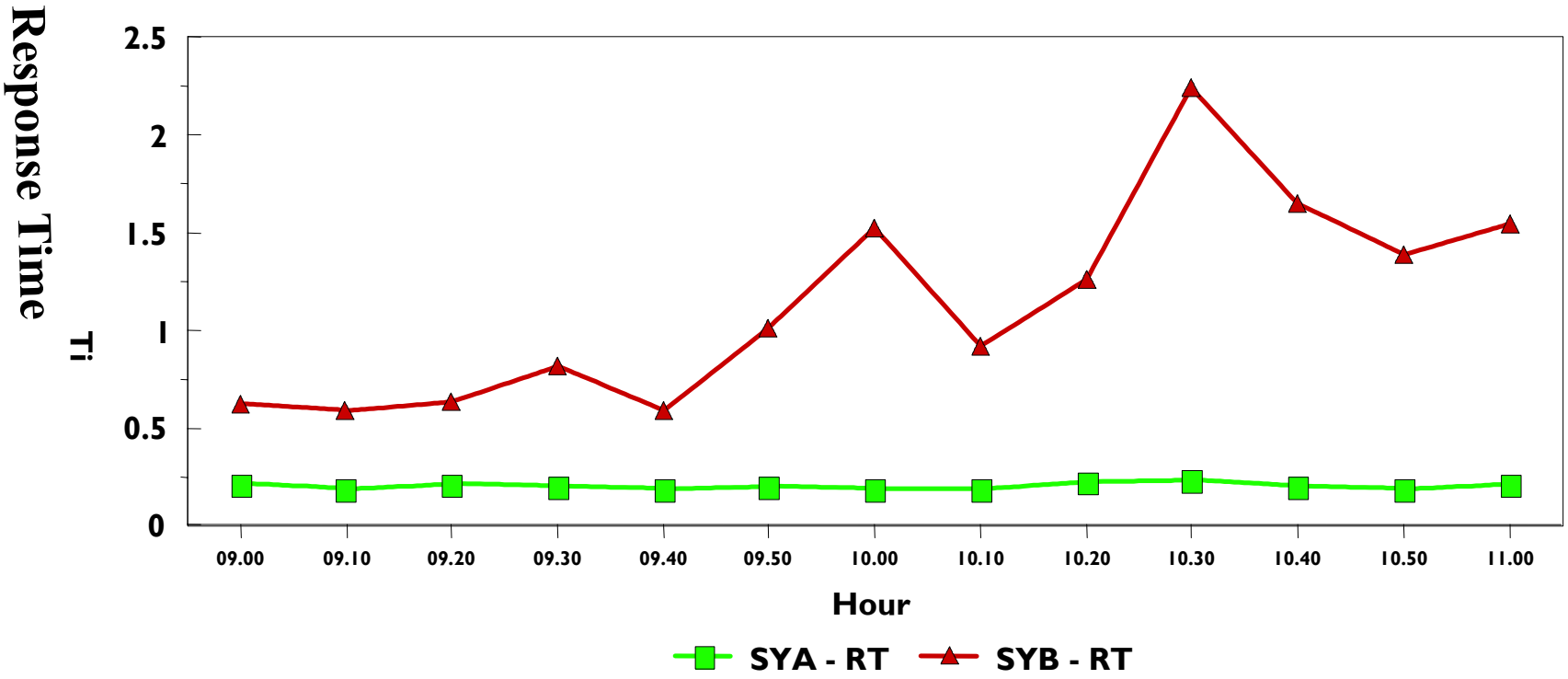
MVS Dispatcher when LPAR Weights are being Enforced

Interval	CP 0	CP 1	CP 2	CP 3
1 CICS,STC,Batch,Batch	CICS L=P	BATCH L=P	STC L=P	BATCH L=P
2 CICS,STC,Batch	CICS L	BATCH L=P	STC L=P	0
3 CICS,Batch,Batch,Batch	CICS L=P	BATCH L=P	BATCH L=P	BATCH L=P
4 CICS	CICS L	0	0	0

```
CICS Active          4:4   = 100%
CICS Dispatched      2:4   =  50%
LPAR BUSY           10:16  =  63%
MVS BUSY            12:16  =  75%
```

Short CP Example

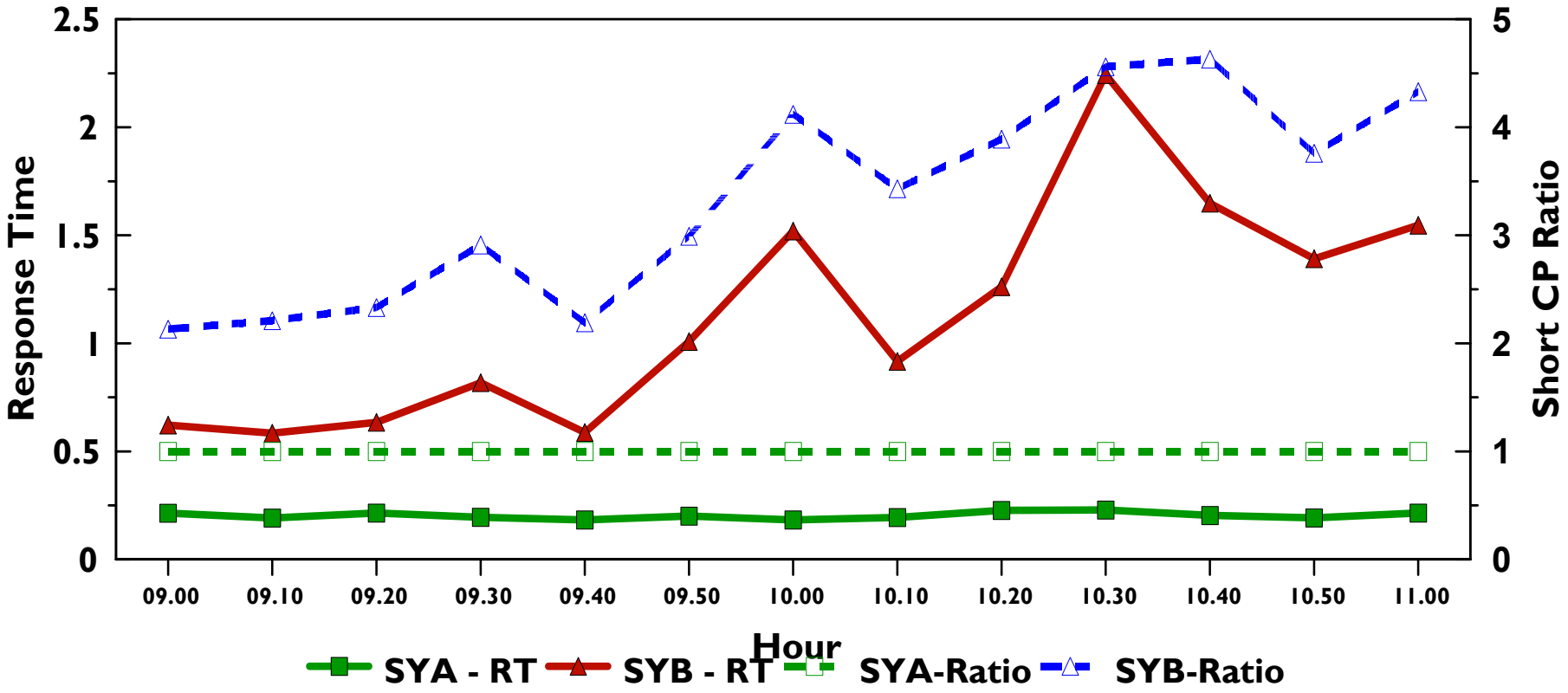
- Data sharing workload runs on 2 images on different CECs
  - Both CECs are the same technology
  - Both LPARs run the same, exact transactions
- Response time on SYB is consistently higher than SYA
- Work is defined with high importance, and a stringent goal



Short CPs - An Example

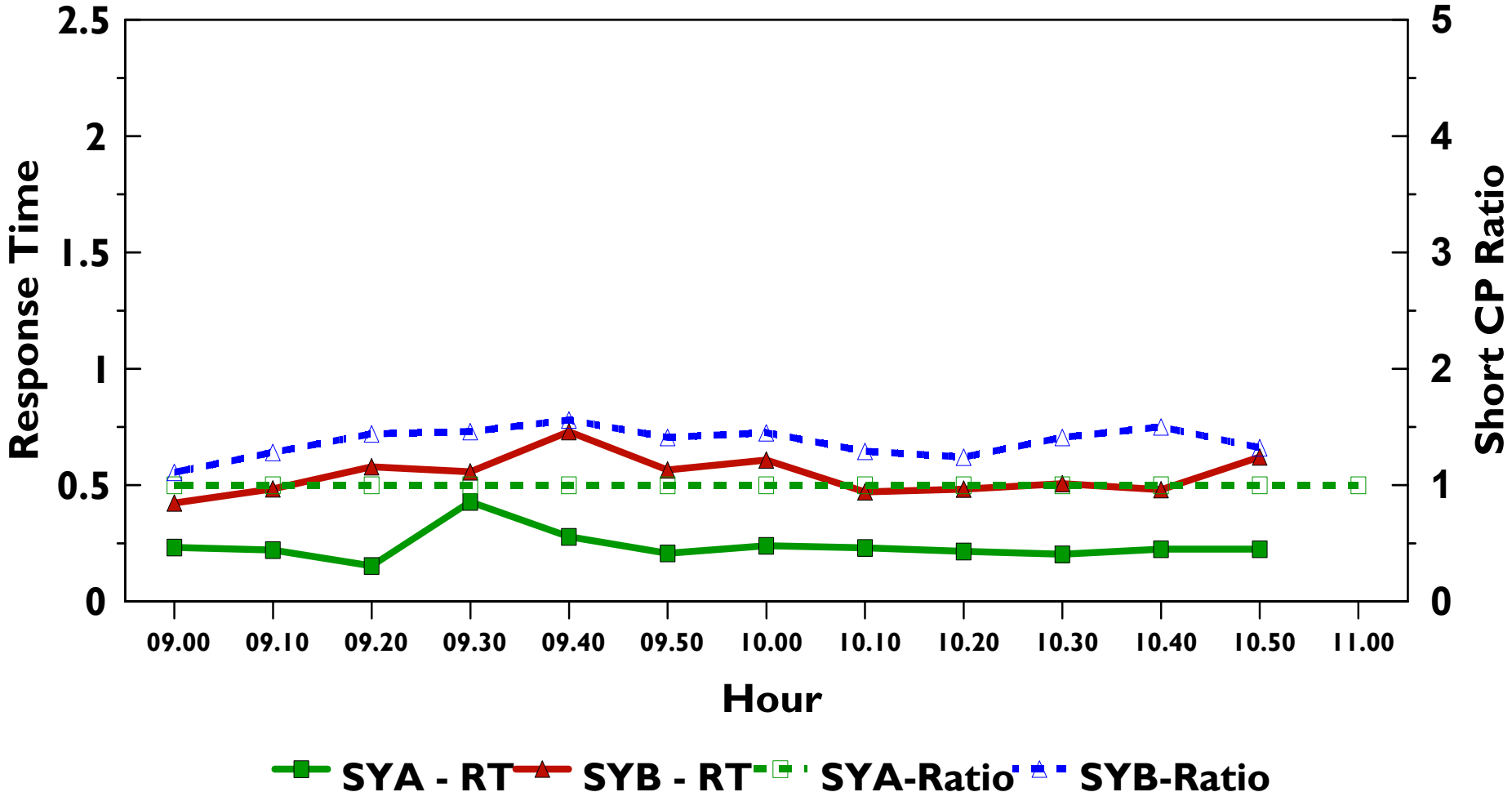
- SYB has short CPs
  - **RMF Short CP ratio = (MVS Busy / LPAR BUSY)**
  - **Each LCP was allowed 18% of a CP across 6 CPs**
  - **Change logicals to 2, get per CP share to 54%**

```
                C P U   A C T I V I T Y
     CPU  2084    MODEL  315
CPU   LPAR BUSY   MVS BUSY
NUM   TIME PERC   TIME PERC
 0    33.39        63.56
 1    31.20        61.34
 3    20.67        50.73
 4    18.26        48.32
CP    TOTAL/AVERAGE   25.88   55.98
```



Short CPs

- Short CP Ratio dropped to approx 1.5 and response times dropped noticeably

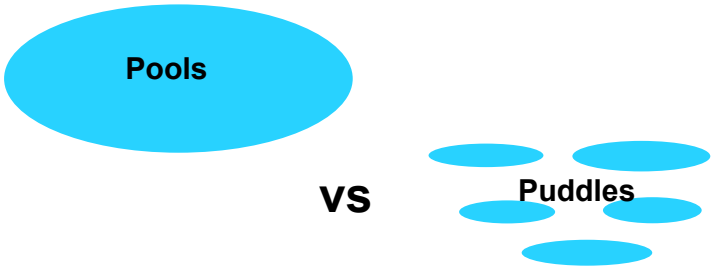


MVS Dispatcher

- Much more functional dispatcher than PR/SM
- Reduced preemption dispatcher
  - Newly ready work at a higher priority is not immediately dispatched
  - Give a small time slice of capacity to newly dispatched work to ensure productive use of the CPU cost to run the dispatcher logic
  - Work at equal dispatch priority use a round robin access to a CP

- If work at a lower dispatch priority is using CPU then higher priority work didn't demand the CP
  - Question which is difficult to answer is: did the work not want the CP (idle) or was the work unable to request the CP (wait)

- Configure systems so more workloads are under the control of the more functional dispatcher
  - WLM is sysplex aware
  - WLM IRD is multiple LPAR aware



Fewer, Faster CPs vs More, Slower CPs

- **Fewer, Faster CPs**
  - High priority workloads see great benefits
  - Have the ability to monopolize a CP
  - On a migration a previously limited workload can now use more capacity
    - Rejoice
    - Control with WLM resource groups
  - Availability Issues

- **More, Slower CPs**
  - More work units are active
  - Can limit a tasks throughput
  - Increased parallelism
  - Limits the impact of a workload which monopolizes a CP

Fewer/Faster Case Study

Case Study Objectives

- New technology is causing many customers to run more partitions on processors with fewer physical CPs
  - Can performance be maintained as the logical to physical ratio increases?
    - What about "short CPs"?
  - What about the overhead of many LPARs on a single machine?
- Evaluate this new environment
- Identify any new performance/capacity planning considerations

Hardware and Software Configurations

Base Configuration

- Software
  - OS/390® V2R10
    - z/OS not supported on 9672-RX3
  - CICS/TS V1R2
  - TPNS
  - Batch

- Hardware
  - Single Partition on a 9672-RX3
    - 1 dedicated CP

9672-RX3

Base Partition OS/390 V2R10 TPNS CICS-TS Batch 1 Dedicated CP	Lightly Used Partition 9 Shared CPs	Lightly Used Partition 9 Shared CPs	Lightly Used Partition 2 Shared CPs	Lightly Used Partition 2 Shared CPs

Migrated Configuration

- Software
  - Same as base configuration
- Hardware
  - Use PCR to estimate capacity of a 2064-116 CP
    - 1 Shared CP

```
PCR (Version 2.5a) - Processor Table, View is Selected, Vendor claim processors are
Excluded Single-CP ITR Ratios relative to IBM 9672-RX3 rated at 1.000
OS/390 (V2R10) - LSPR Rel 2002a (02/13/2002)
                                          Custom   20%      20%      20%      20%      20%
 Processor    Features    F SG   MSU       Mix     CBW2     CB84     TSO    CICS/DB2  IMS
 ---------- ---------- - --   ---     ------   ------   ------   ------   ------   ------

 9672-RX3     10W          70   30      1.000    1.000    1.000    1.000    1.000    1.000
 2064-116     20PU 16W        441      9.303   10.772    9.692    8.322    8.316    9.899
```

2064-116

RAP01	RAP02	RAP03	RAP04	RAP05	RAP06	RAP07	RAP08	RAP09	RAP10	SOAKER
1 Shared CP	1 Shared CP	1 Shared CP	1 Shared CP	1 Shared CP	1 Shared CP	1 Shared CP	1 Shared CP	1 Shared CP	1 Shared CP	15 Ded. CPs

Workload Characteristics

Workloads

- TPNS
  - SYSSTC service class
  - Simulate a 500 terminal network
  - Vary think time to drive 9672-RX3 partition to greater than 80% busy
    - 5-second think time
- CICS-TS
  - Single CICS region
  - Multiple transactions accessing VSAM files
  - Transaction classification
    - Average response time LE .125 seconds
  - Run during the entire measurement period
- Batch
  - Submitted back end of the measurement
    - Push partition to 100% busy to cause LPAR weights to be enforced
      - 1 address space

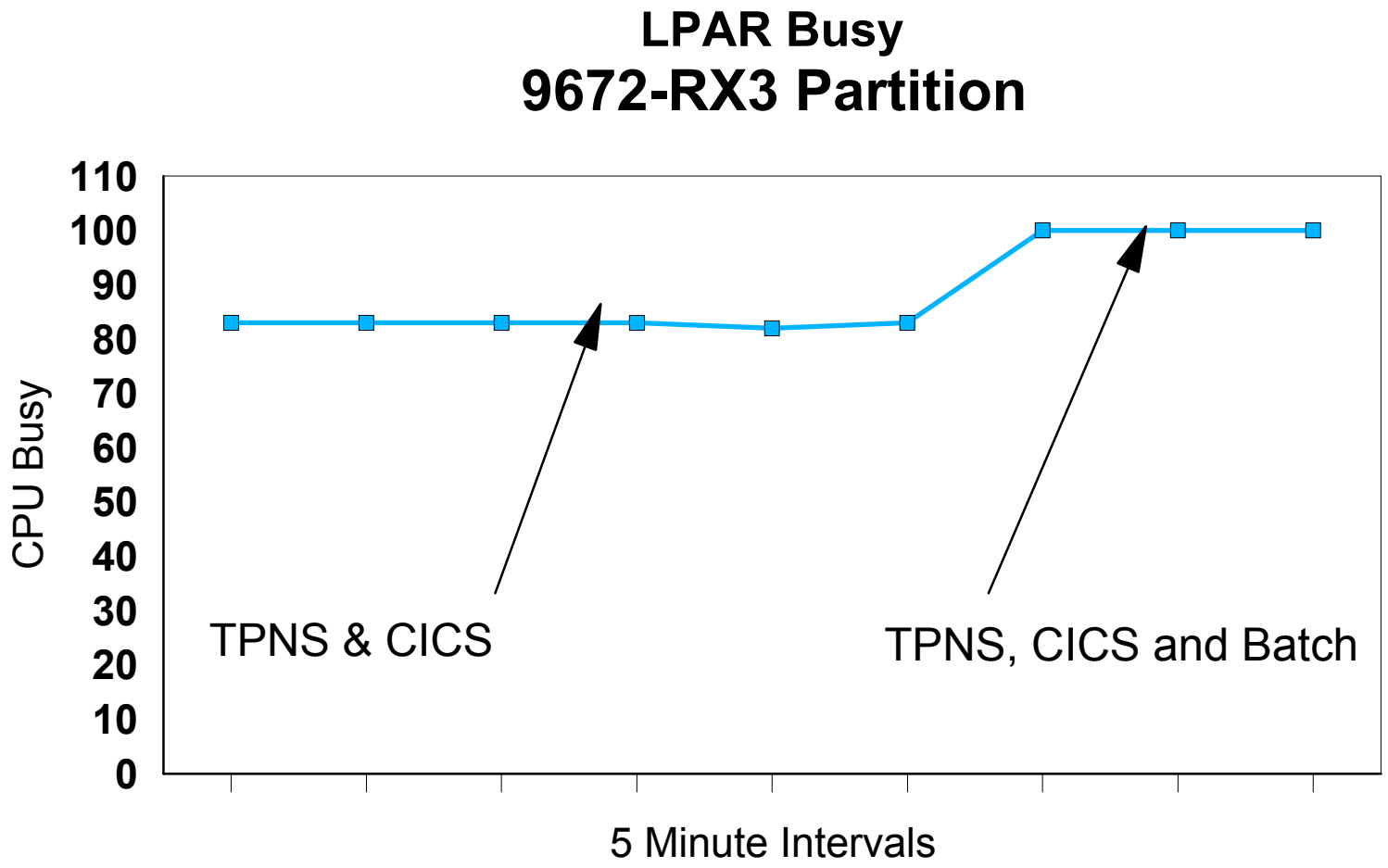
Monitored Measurement Variables

RMF Metrics

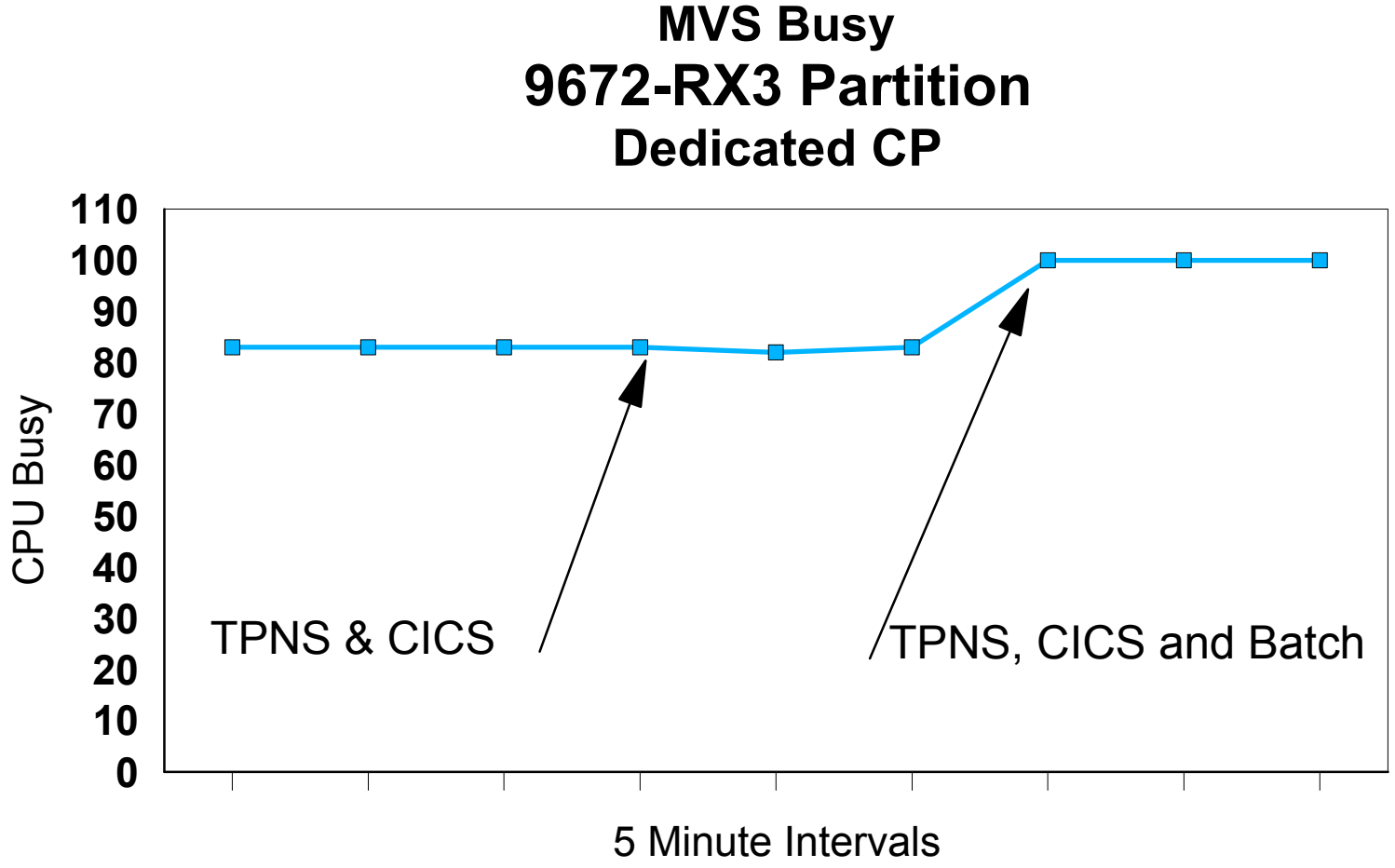
- MVS Busy
  - How busy is the partition from an MVS point of view?
  - Indicator of "short CPs" (LPAR taking the physical CP away from the logical CP)
- LPAR Busy
  - How busy are the logical CPs in the partition?
    - How busy is the base partition? (9672-RX3)
    - How busy are each of the migrated partitions? (2064-116)
- CICS Transaction Values
  - Rate
    - How many answers is the CICS system generating?
  - Response Time
    - How long does it take to make an answer?
  - Performance Index
    - How well did the transaction achieve its goal?

Test Results

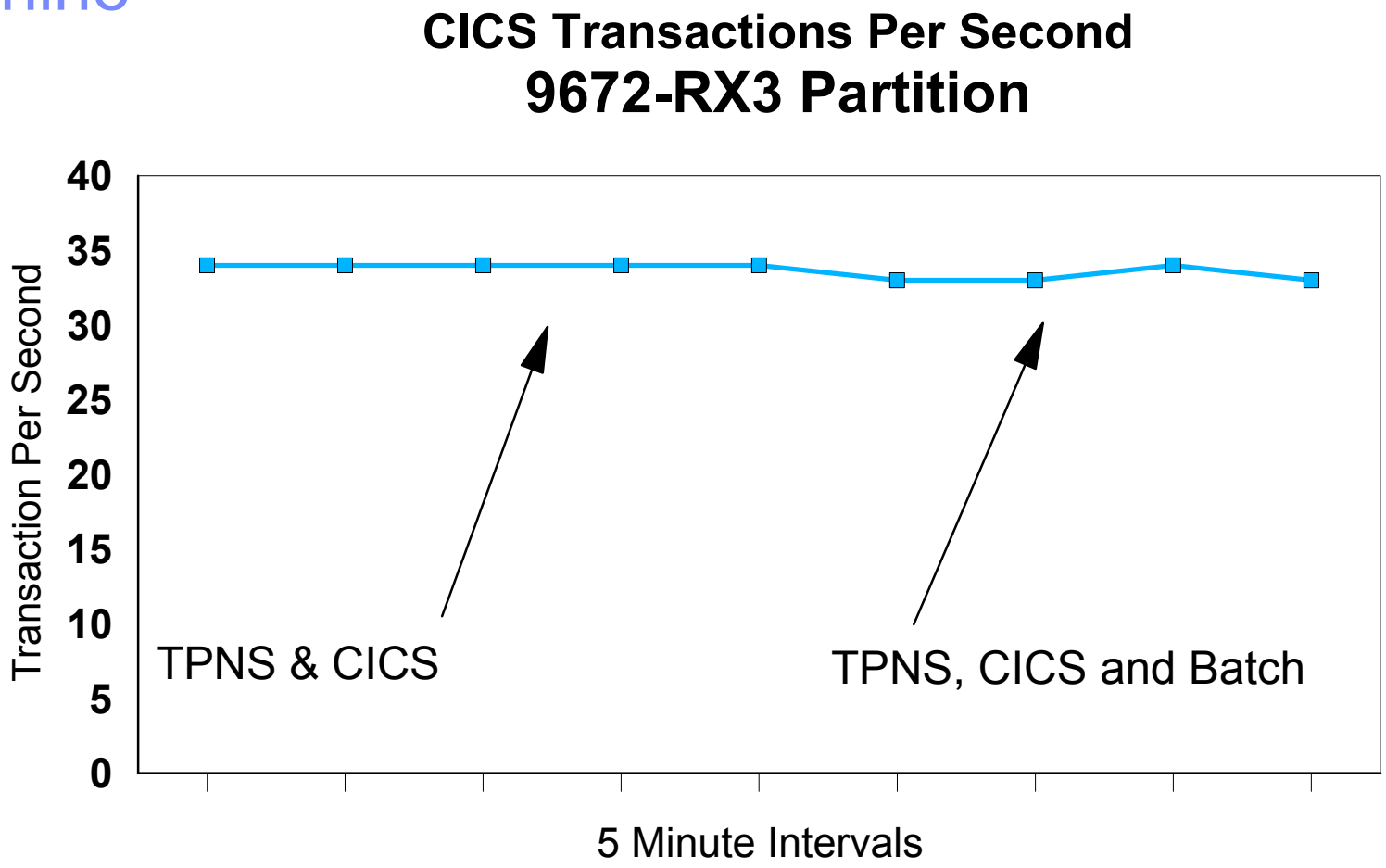
Base Machine



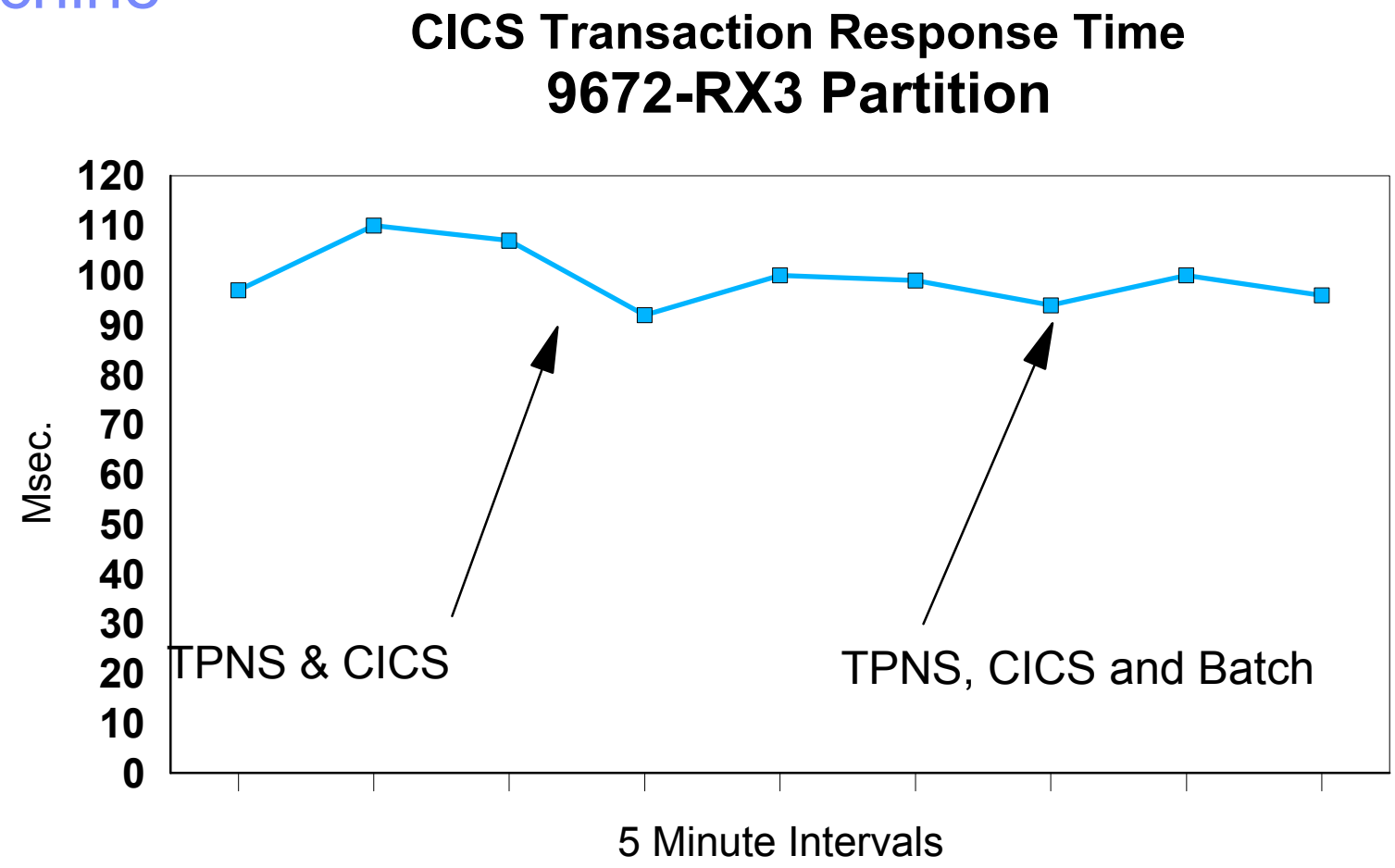
Base Machine



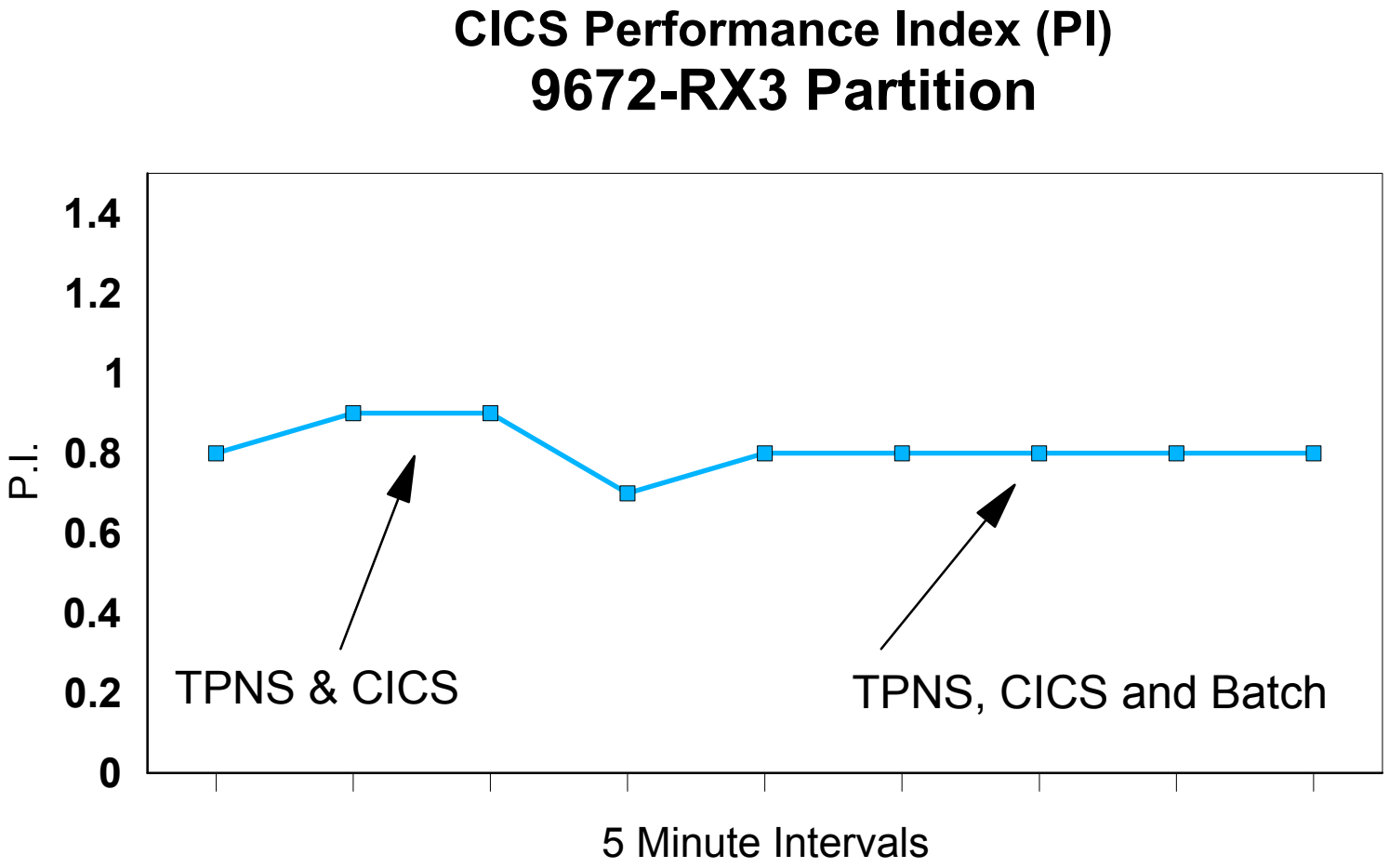
Base Machine



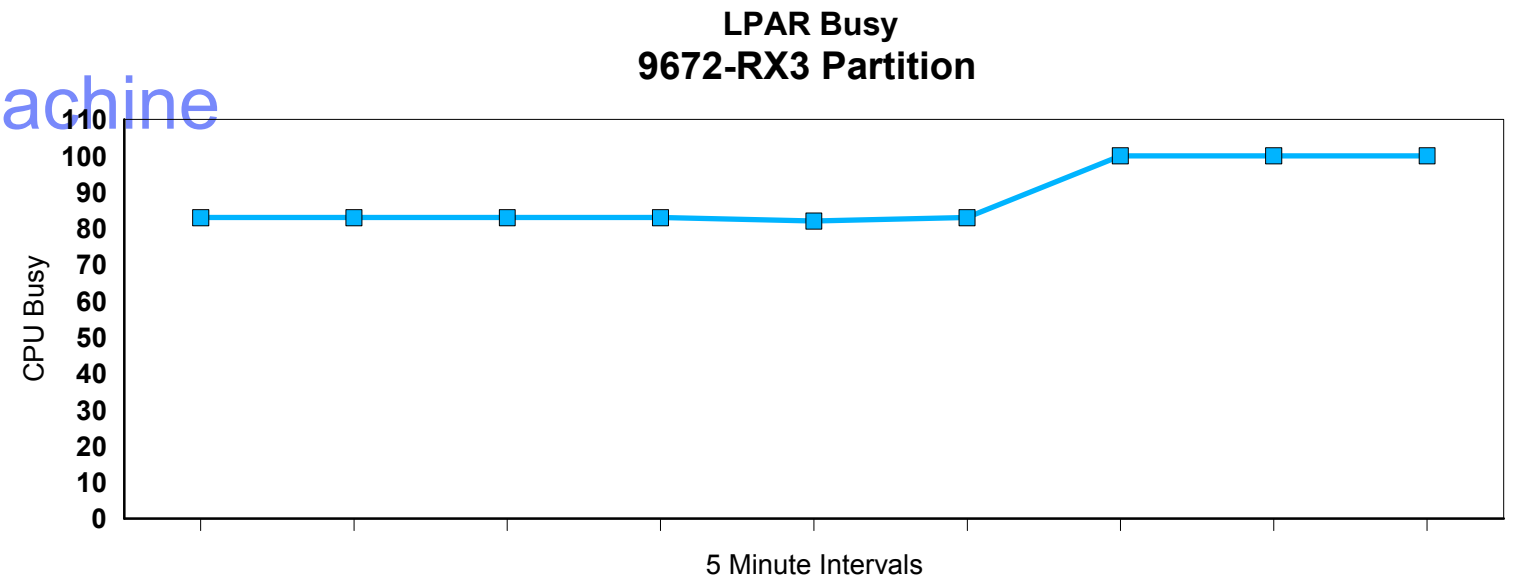
Base Machine

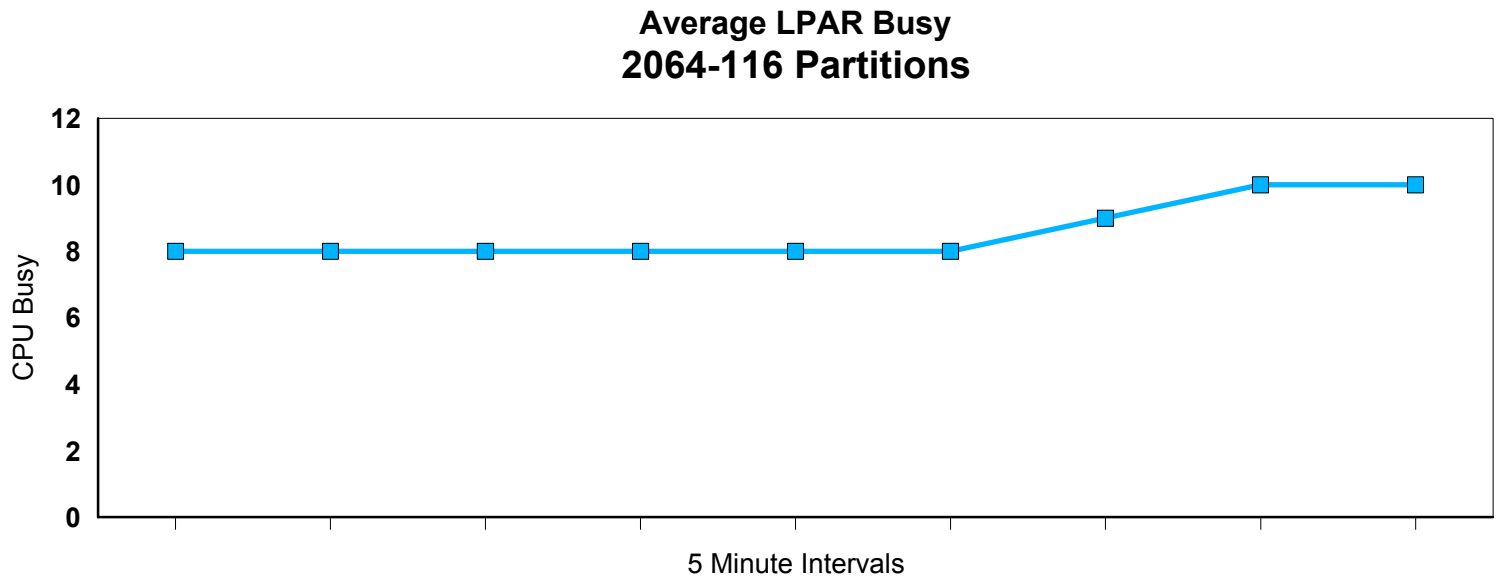


Base Machine

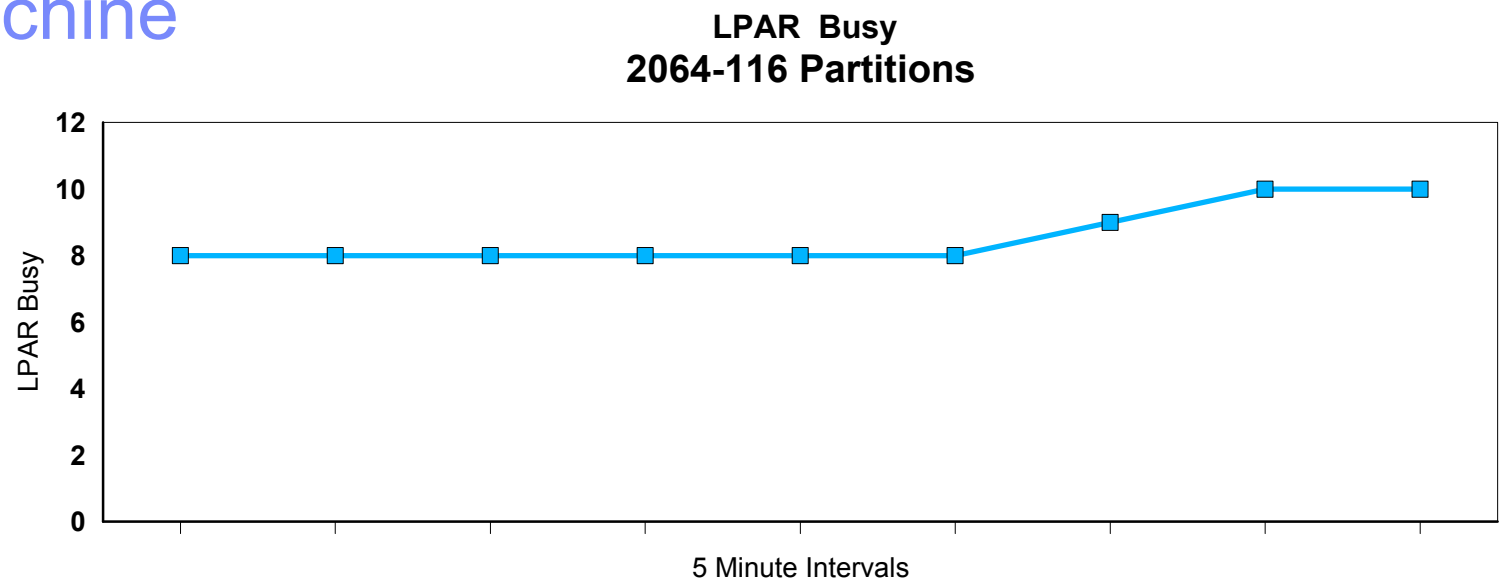


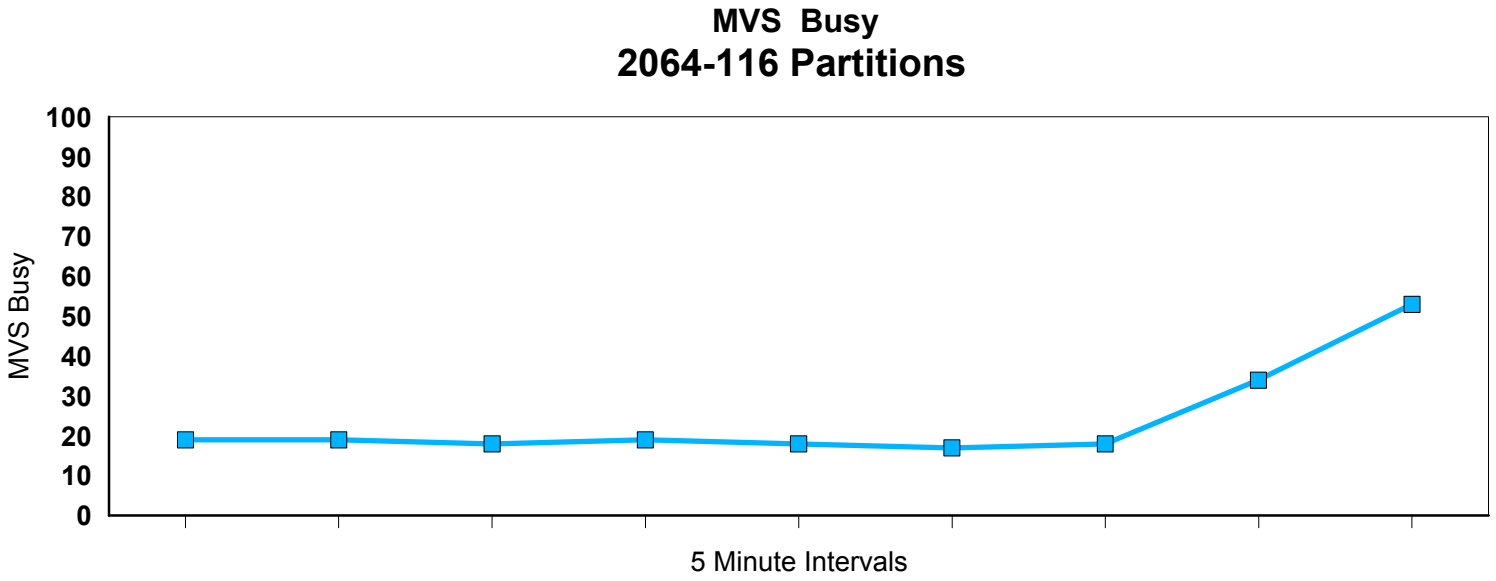
Migrated Machine



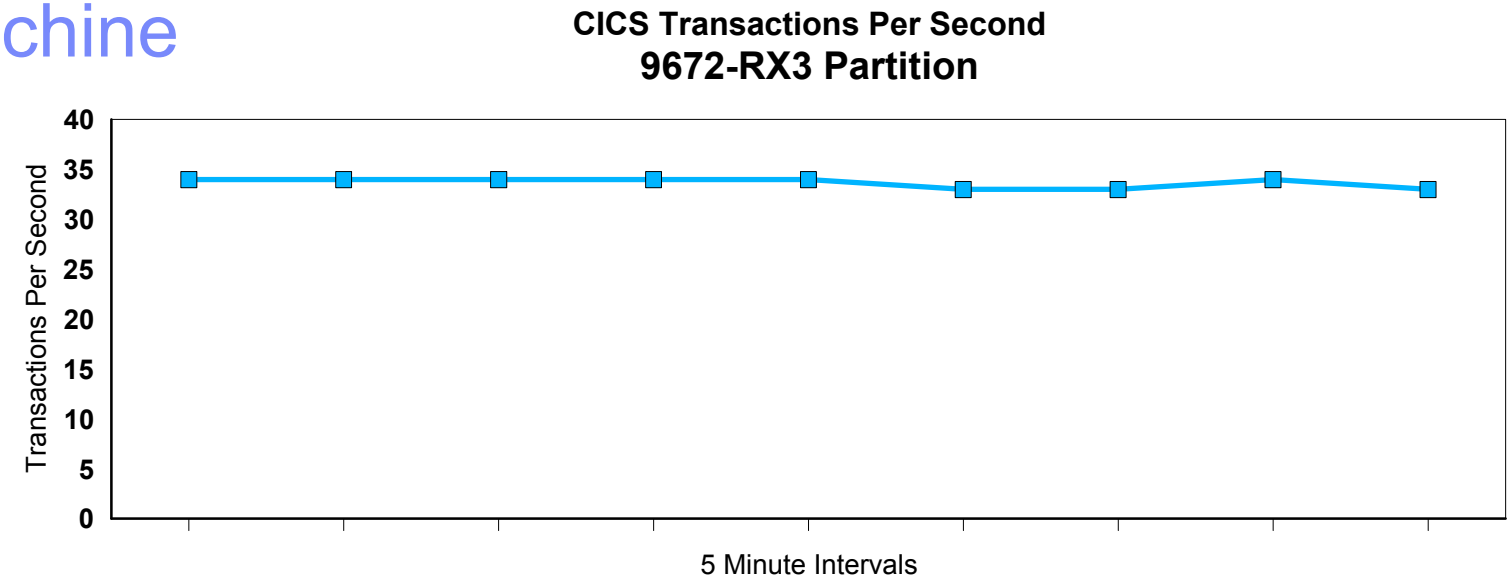


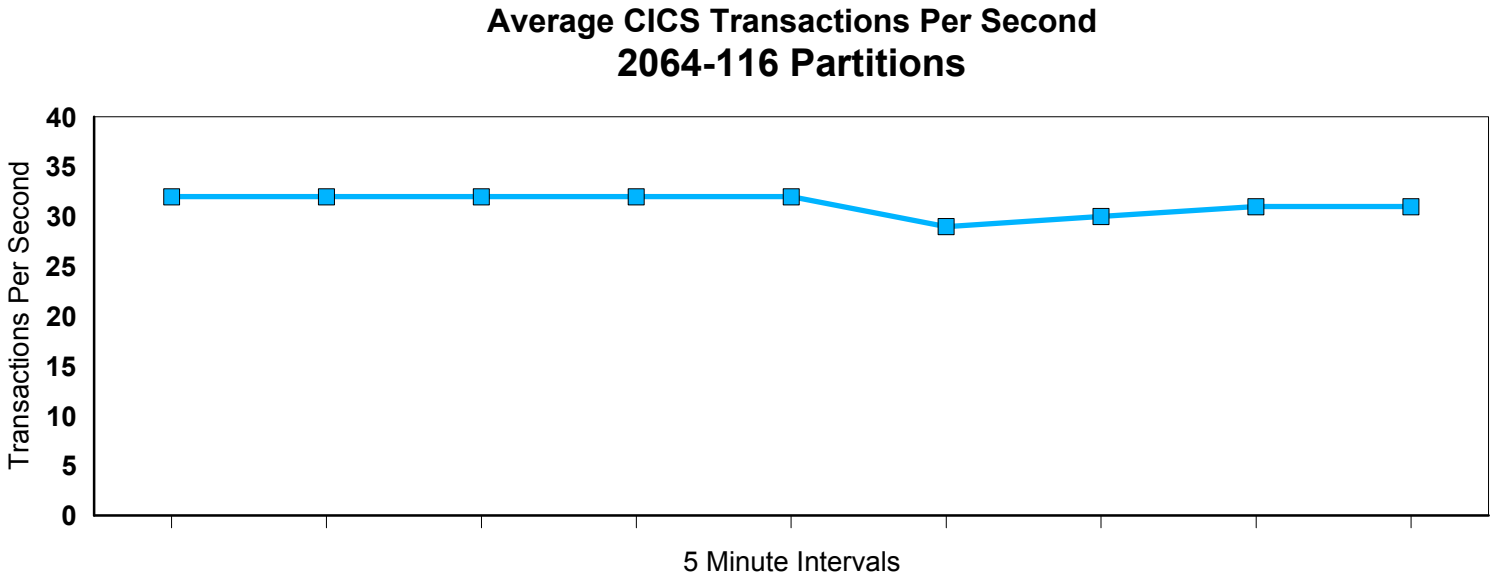
Migrated Machine



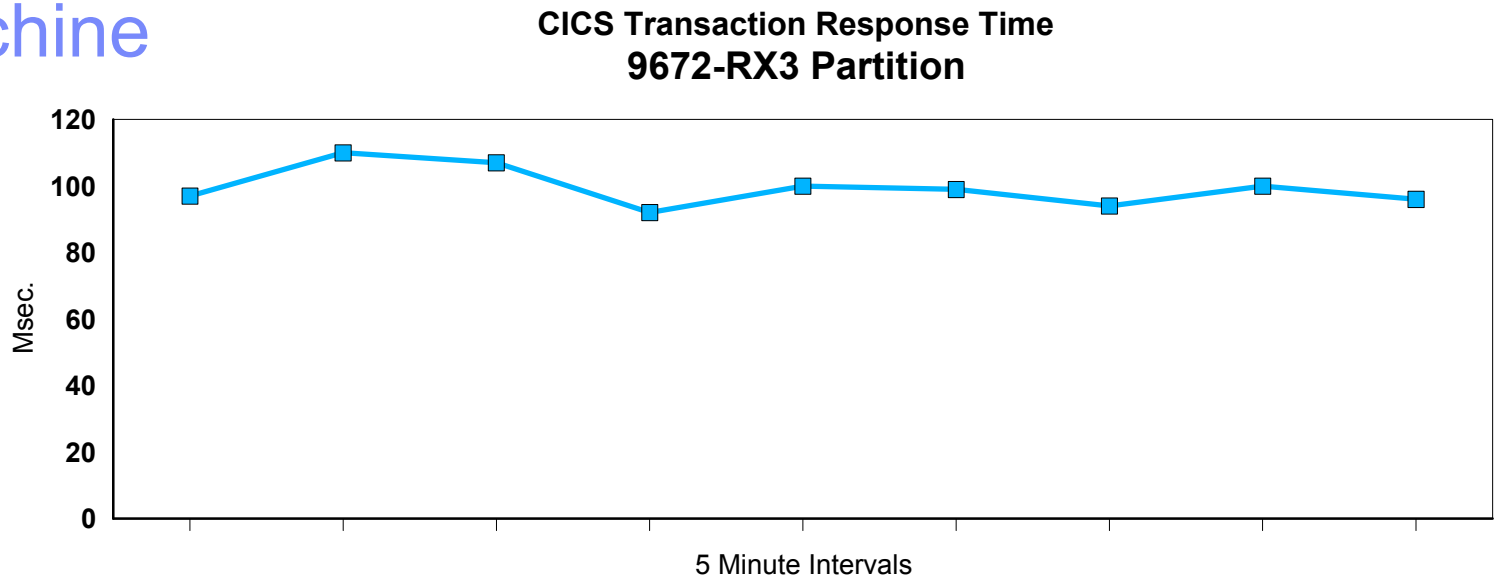


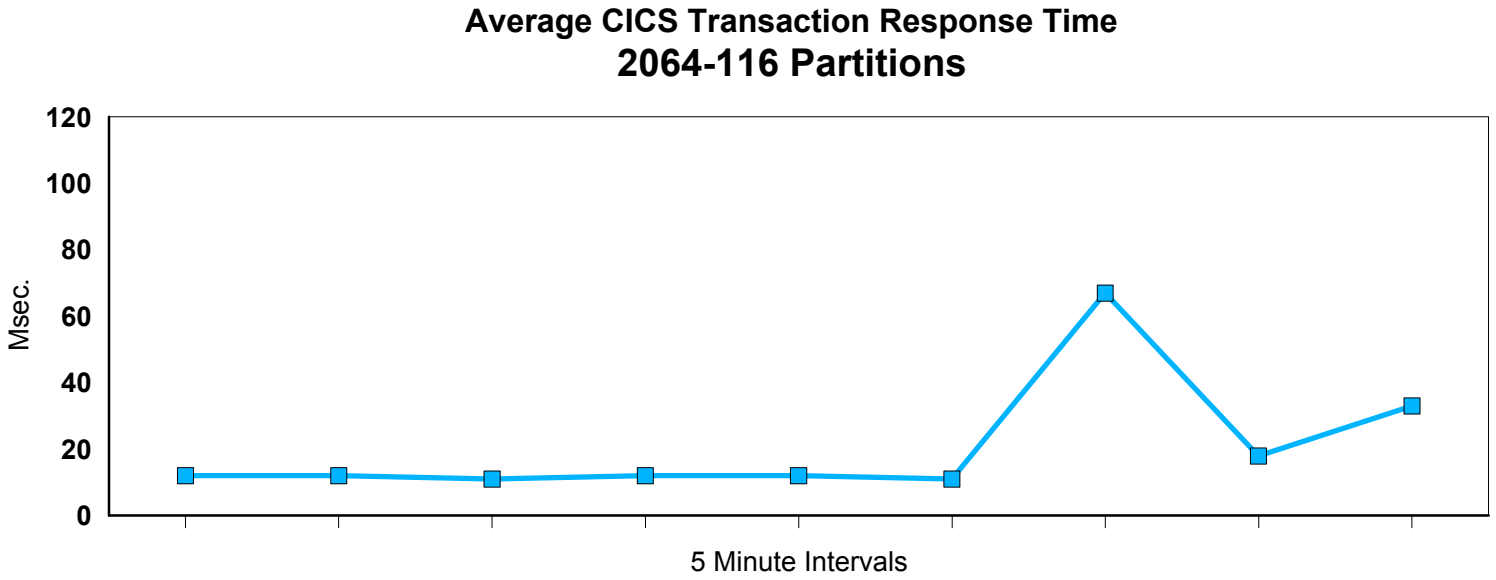
Migrated Machine



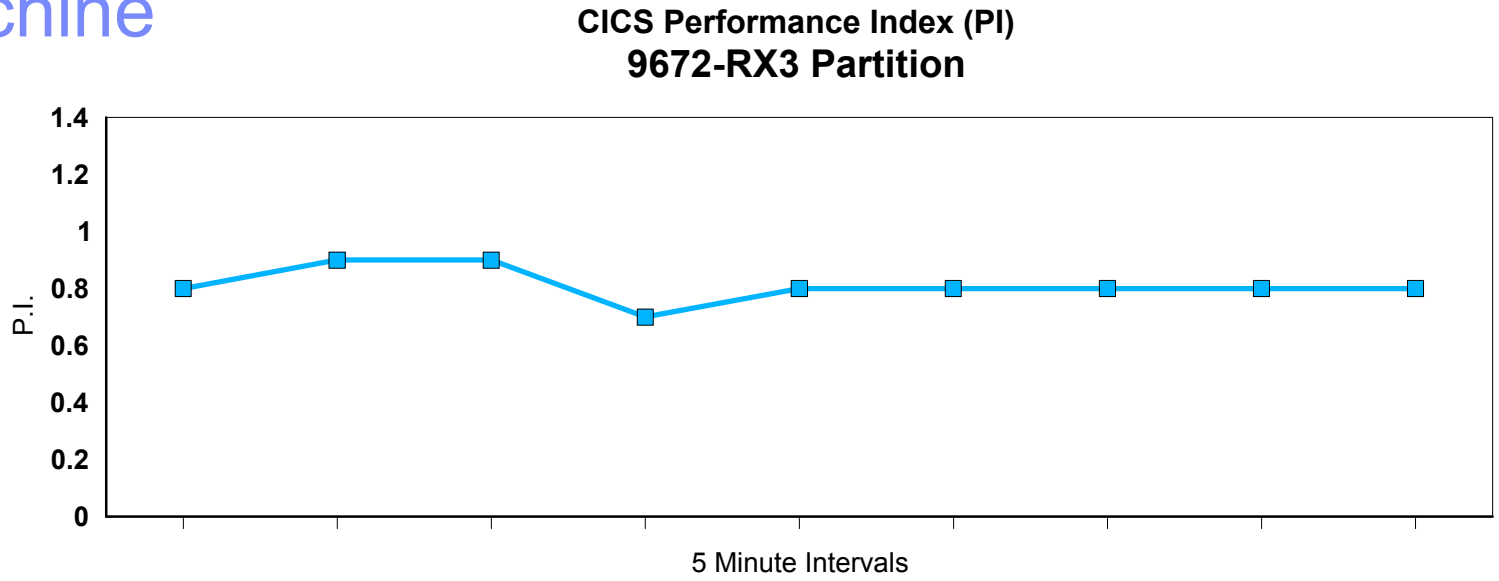


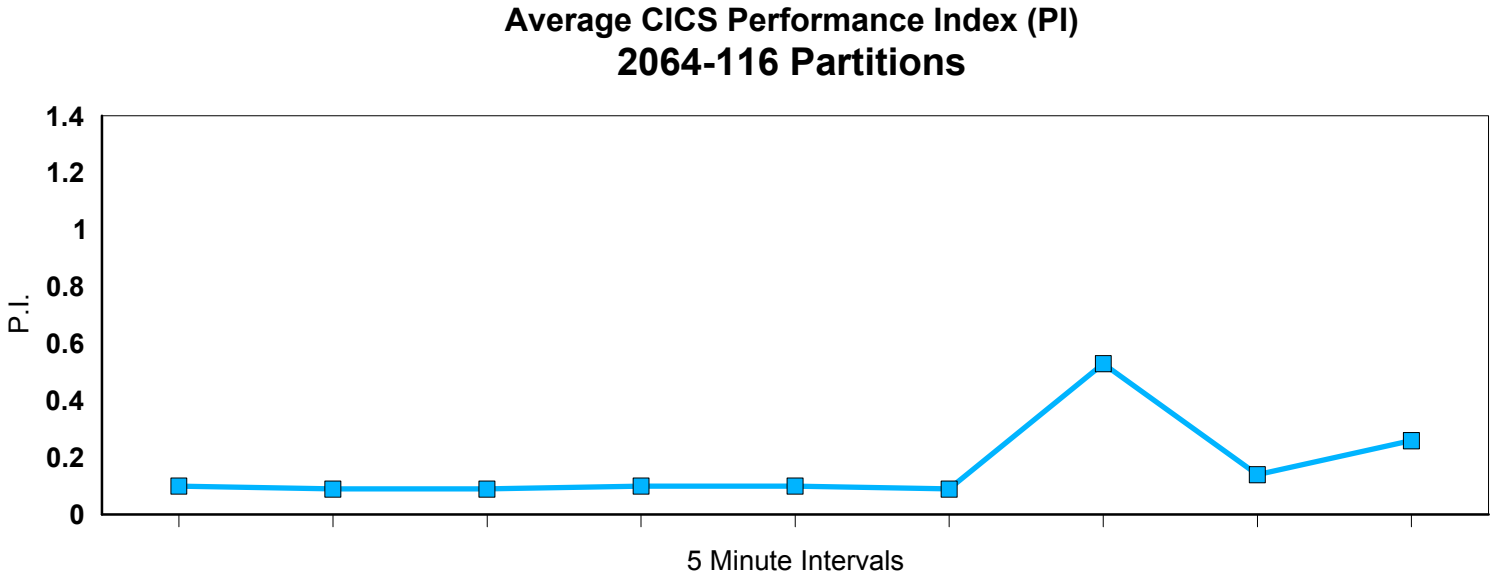
Migrated Machine





Migrated Machine





Recommendations

How to be Successful

- Don't Ignore "Short CPs"; to be successful, make sure each:
  - Partition has access to the same or more MIPS on the new machine
  - Logical CP has access to the same or more MIPS on the new machine
- The overhead of LPAR can make a difference when the logical to physical CP ratio is greater than 3-to-1
  - Use LPAR/CE or IBM Processor Capacity Reference (zPCR) to estimate LPAR cost

```
PCR (Version 2.5a) - Processor Table, View is Selected, Vendor claim processors are
Excluded Single-CP ITR Ratios relative to IBM 9672-RX3 rated at 1.000
OS/390 (V2R10) - LSPR™ Rel 2002a (02/13/2002)
```

Processor	Features	F	SG	MSU	Custom Mix	20% CBW2	20% CB84	20% TSO	20% CICS/DB2™	20% IMS™
9672-RX3	10W		70	30	1.000	1.000	1.000	1.000	1.000	1.000
2064-116	20PU 16W			441	9.303	10.772	9.692	8.322	8.316	9.899

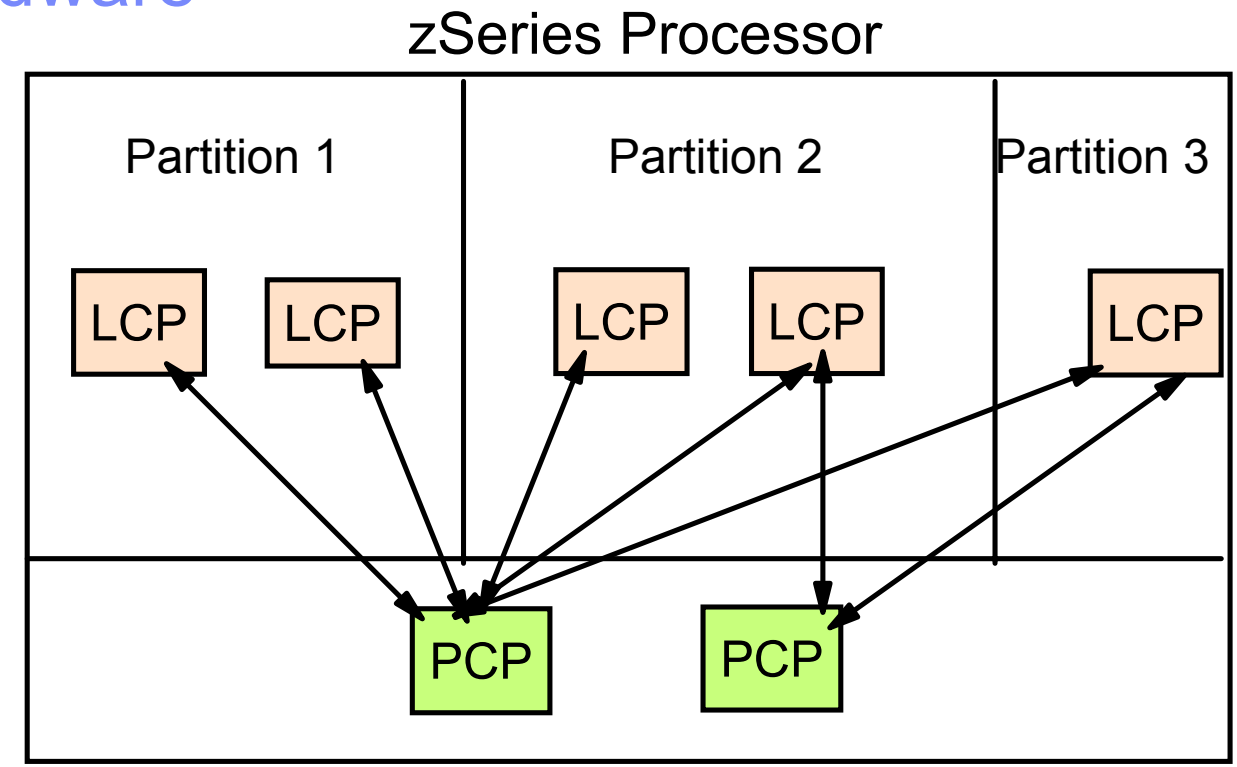
IBM CPS
Capacity Planning Support
LPAR Capacity Estimator

Estimate of host's LPAR-mode (vs.) host's B-mode capacity: **80.62%**

Causes of LPAR Overhead

Management of the Hardware

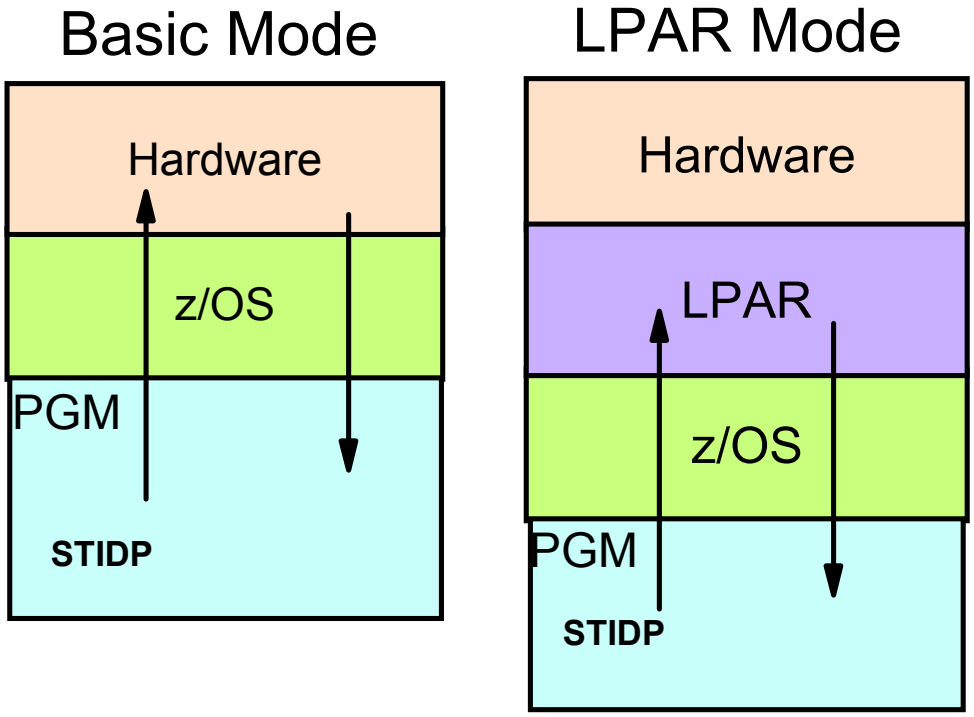
- **LPAR must assign a physical CP to a logical CP in order to execute instructions**
- **CPU cost in the *PHYSICAL partition on RMF Partition Report**
- **CPU Timers are stopped when physical CP is removed from the logical CP**



Logical to Physical Ratio	LPAR Overhead
1:1	0.22%
2:1	0.42%
3:1	0.55%
10:1	0.81%

Management of the Partition

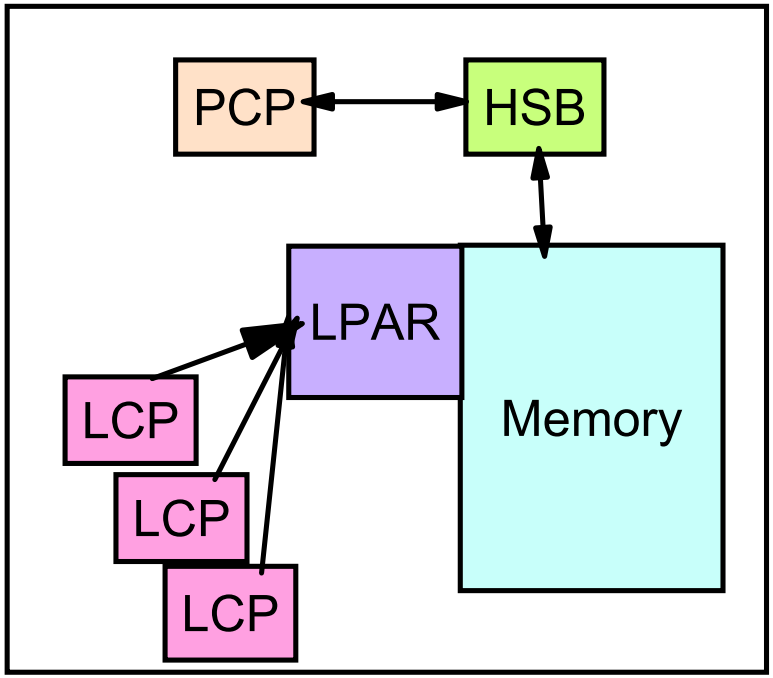
- **LPAR must get involved with certain instructions**
- **Cost reflected in LPAR MGMT Column on RMF Partition Report for each partition**




```
--------- PARTITION DATA --- AVERAGE PROCESSOR UTILIZATION PERCENTAGES --
                          LOGICAL PROCESSORS  --- PHYSICAL PROCESSORS ---
NAME       S   WGT  DEF   EFFECTIVE    TOTAL  LPAR MGMT  EFFECTIVE  TOTAL
SYS6LP01   A   174    0       85.29    85.53       0.24      85.29  85.53
SYS6LP02   A    35    0       53.86    55.90       0.51      13.46  13.97
*PHYSICAL*                                         0.44                0.44
                                                  -----      ----- ------
  TOTAL                                            1.19      98.75  99.94
```

High Speed Buffer Contention

- **The high speed buffer is 'fast' memory**
  - **Accessing data from the high speed buffer improves the speed of the PCP**
  - **Data not found in the high speed buffer reduces the effective speed of the PCP**
- **Each time a new LCP is associated with a PCP, increased risk of HSB miss**
- **The impact of the HSB miss is not reported in RMF, but is reflected in increased TCB time for jobs**
- **IBM tool zPCR includes estimated TCB time elongation**



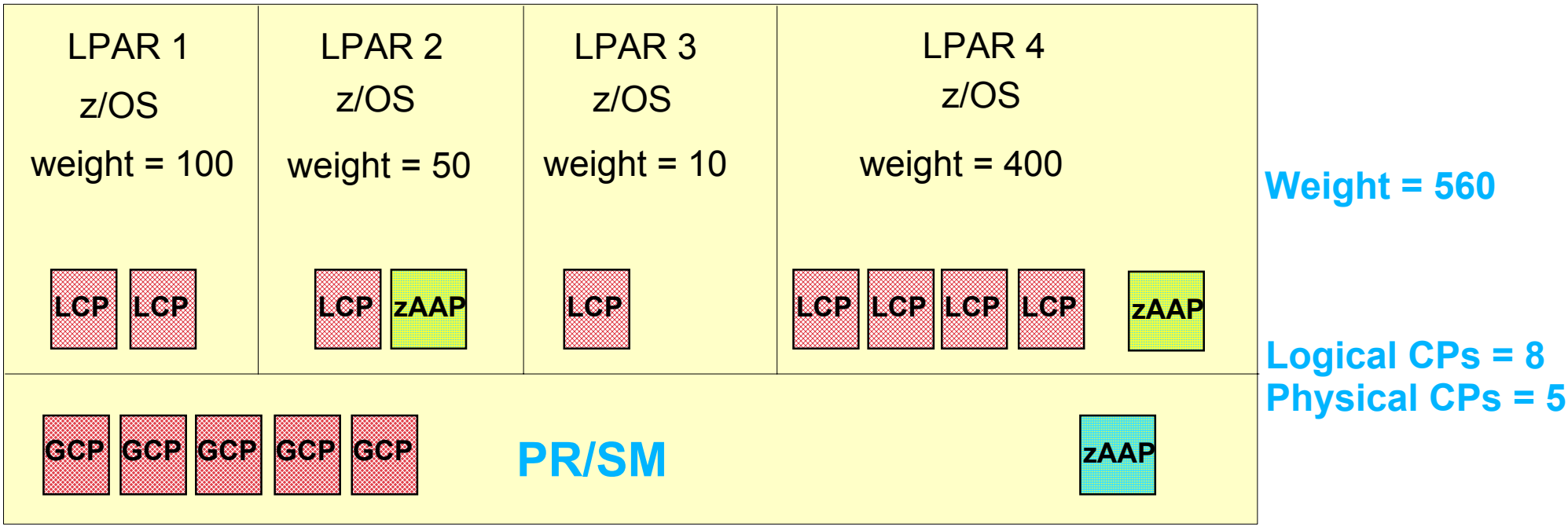
Miscellaneous LPAR Information

Logical to Physical CP Ratio

- Strive to keep logical to physical ratio in the 2:1 or 3:1 area
- Higher ratios will work but cause increased cost which needs to be factored into the capacity plan
- Biggest issue to reducing the logical to physical CP ratio is the requirement to run small LPARs as z/OS uni-processors
  - Availability issues of running z/OS as a uni-processor
  - Places greater emphasis on doing LPAR consolidation to make fewer LPARs which need more than 1 CP of capacity
    - Virtual storage constraints need to be reviewed
  - Places greater emphasis on doing CICS consolidation to make fewer, larger CICS regions which can use more of a CP's capacity
    - Virtual storage constraints need to be reviewed

Capacity Planning and LPAR

- **n-way and MP effects will impact capacity**
- **LPAR 3 is a uni, but the hardware is running as a 6-way shared processor and the capacity is of a 6-way shared processor**
  - **5 GCPs and 1 zAAP**
  - **z/OS 1.6 will support up to 24 CPs per image**



Capacity Planning and LPAR

- **Set a 2084-316 as the base processor equated to 1.0**
  - 5752 MIPS, and defined with a LowIO mixed workload and shared CPs

Case	Mode	# of LPARs	LPARs x LCPs	LCP	ITRR	LCP:PCP
Base	2084-316	1	1 x 16	16	1.00	1:1
1	2084-316	2	2 x 12	24	.9882	1.5:1
2	2084-316	7	2 x 12 2 x 4 3 x 2	38	.9591	2.4:1
3	2084-316	4	2 x 12 2 x 6	36	.9733	2.25:1
4	2084-316	4	4 x 6	24	.9998	1.5:1