

# Route me, Workload Manager

WLM functions for  
dynamic workload routing

Horst Sinram, STSM, z/OS Workload and Capacity Management, [sinram@de.ibm.com](mailto:sinram@de.ibm.com)  
IBM Germany Research & Development 19 Jul 2018

# Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

AIX*	DS8000*	Language Environment*	SystemPac*	z10
BladeCenter*	FICON*	Parallel Sysplex*	System Storage	z10 BC
DataPower*	HiperSockets	POWER7*	System z	z10 EC
DB2*	Hyperwap	PrintWay	System z9	z/OS*
DFSMS	IBM*	ProductPac*	System z10	zEnterprise
DFSMSdss	IBM eServer	RACF*	System z10 Business Class	zSeries*
DFSMSHsm	IBM logo*	REXX	WebSphere*	
DFSMSrmm	ibm.com	RMF	z9*	
DFSORT	Infiniband*	ServerPac*		
DS6000*	InfoPrint			

\* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

InfiniBand is a registered trademark of the InfiniBand Trade Association (IBTA).

Intel is a trademark of the Intel Corporation in the United States and other countries.

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

Java and all Java-related trademarks and logos are trademarks or registered trademarks of Sun Microsystems, Inc., in the United States and other countries.

Microsoft, Windows and Windows NT are registered trademarks of Microsoft Corporation.

UNIX is a registered trademark of The Open Group in the United States and other countries.

All other products may be trademarks or registered trademarks of their respective companies.

The Open Group is a registered trademark of The Open Group in the US and other countries.

## Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed.

Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved.

Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

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

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This presentation and the claims outlined in it were reviewed for compliance with US law. Adaptations of these claims for use in other geographies must be reviewed by the local country counsel for compliance with local laws.

# Agenda



## ▪ Concepts

- Importance levels
- Displaceable capacity
- Free capacity

## ▪ WLM Sysplex Routing Services

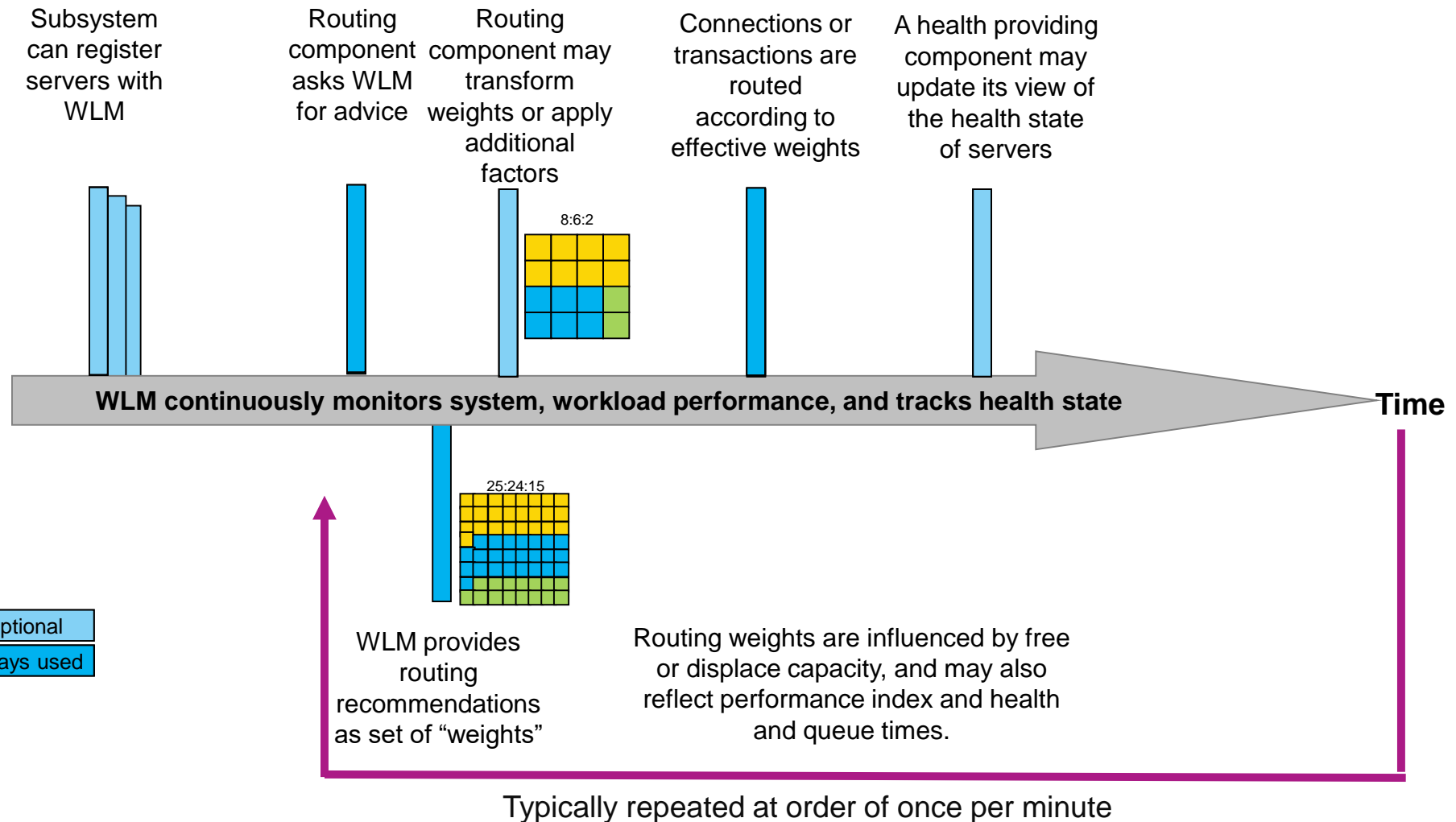
- IWMWSYSQ
- IWMSRSRS
- IWM4SRSC
- Basic capacity-based weights and additional influencers

## ▪ Observations, best practices and optimization approaches

## WLM Dynamic Workload Routing Services

- WLM Sysplex routing services provide guidance to routing components on how to distribute
  - Transactions
  - Connections
- Multiple sets of routing APIs are offered by WLM
  - Same underlying view of “capacity” but different algorithms and influencing parameters
- Scope
  - Multiple systems of one Sysplex, one or more servers per system
- Primary objectives for balancing:
  - Capacity – Route work according to capacity available
  - Performance – WLM goal attainment
  - Availability – Avoid shortages
  - Reliability – Avoid unhealthy work consumers

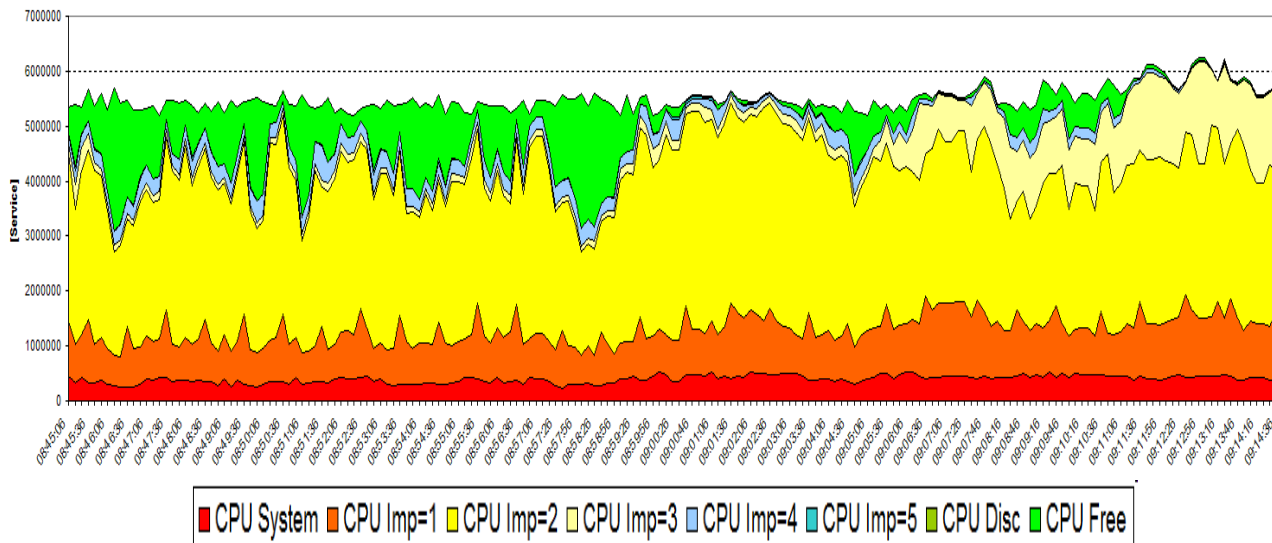
# The life cycle of workload routing recommendations



## Concepts: Service consumption by importance level

- WLM/SRM tracks the consumption of *CPU service* by importance level
- WLM management will sacrifice less important work to allow more important work to achieve its goal. Less important work may even be *displaced* entirely.

CPU Service Consumption by Importance Level



- Level 0:  
SYSTEM and SYSSTC  
■ CPU System
- Level 1-5:  
Importance 1 through 5  
■ CPU Imp=1 ■ CPU Imp=2 ■ CPU Imp=3 ■ CPU Imp=4 ■ CPU Imp=5
- Level 6:  
Discretionary  
■ CPU Disc
- Level 7:  
Free (unused) capacity  
■ CPU Free

## Concepts: WLM determination of displaceable capacity

- An important metric for routing decisions is the *displaceable capacity* at a given importance level (i):

$$DisplaceableCapacity_i = FreeCapacity + \sum_{j=i+1}^6 CapacityConsumed_j$$

or

$$DisplaceableCapacity_i = \sum_{j=i+1}^7 CapacityConsumed_j$$

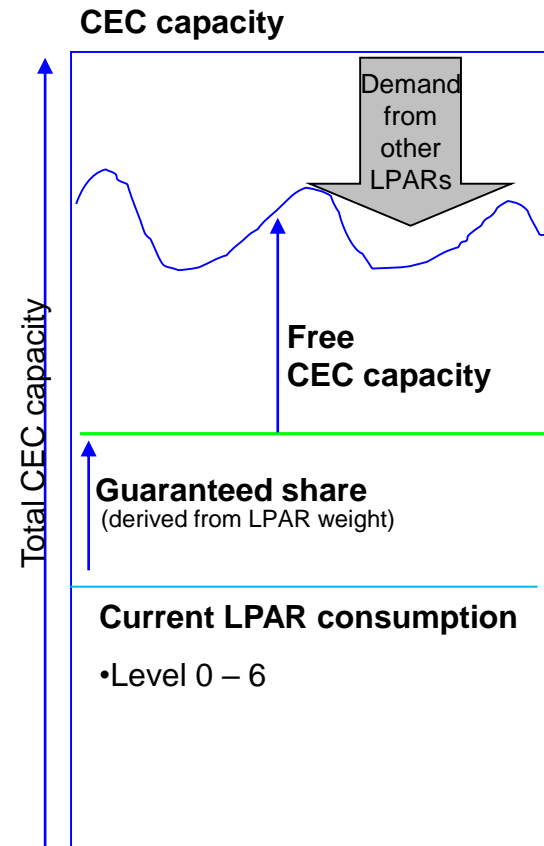
- For the purpose of routing the **3 min rolling averages** of consumption and free capacity are considered
- The *consumed capacity* is usually well understood
  - *Free capacity* may be harder to understand
  - Needs to reflect many different constraints that could limit the capacity that can be consumed by an LPAR.
- All processor types (CP, zIIP, zAAP) to be assessed independently

# Concepts: LPAR Capacity

## What limits the capacity of an LPAR?

- Logical capacity (number of logical processors)
- LPAR weight
  - Guaranteed capacity unless configuration parameters prohibit the guaranteed capacity to be consumed
  - IRD weight management may change weights dynamically hence guaranteed capacity changes
  - Or *dedicated* LPAR
- LPAR initial cap (*hard cap*), LPAR absolute cap (since zEC12 GA2)
- Defined capacity (*soft cap*)
  - LPAR level defined capacity
  - Group capacity

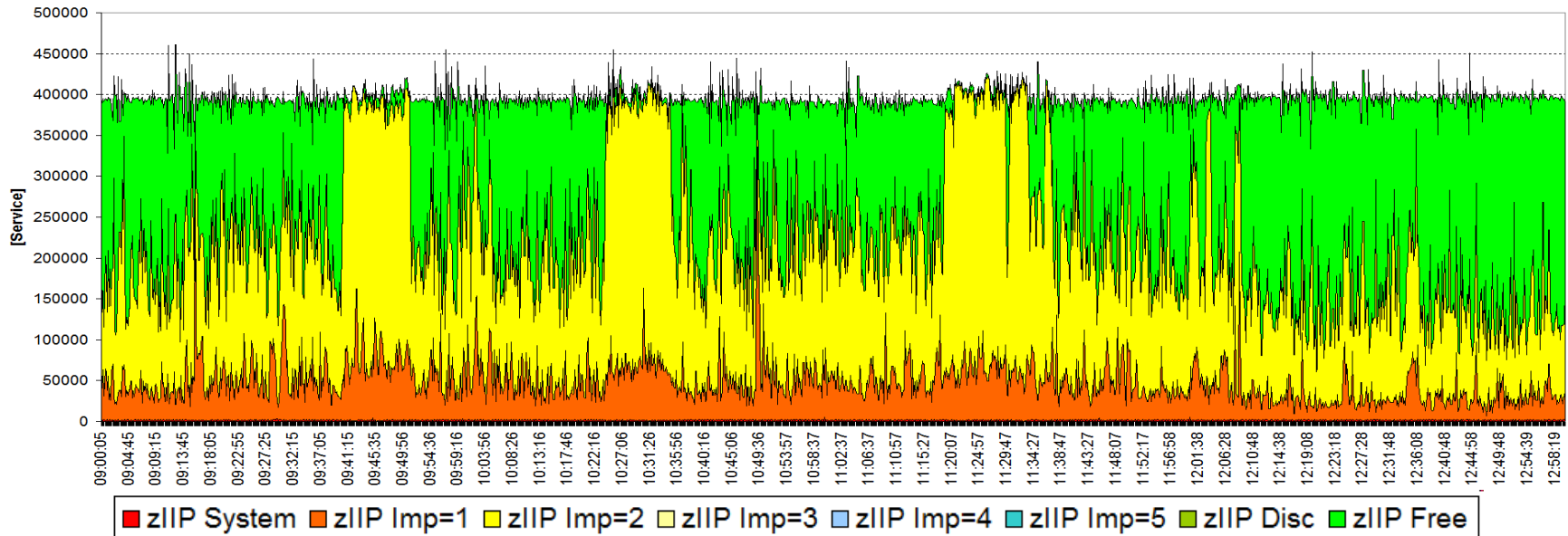
Defined capacity is only considered while the 4HRA consumption exceeds to defined limit , unless AbsMSUCapping is in effect.
- Available CEC capacity – unused CEC capacity can be consumed beyond weight
- In addition, consider
  - MVS Busy (MVS wait time)





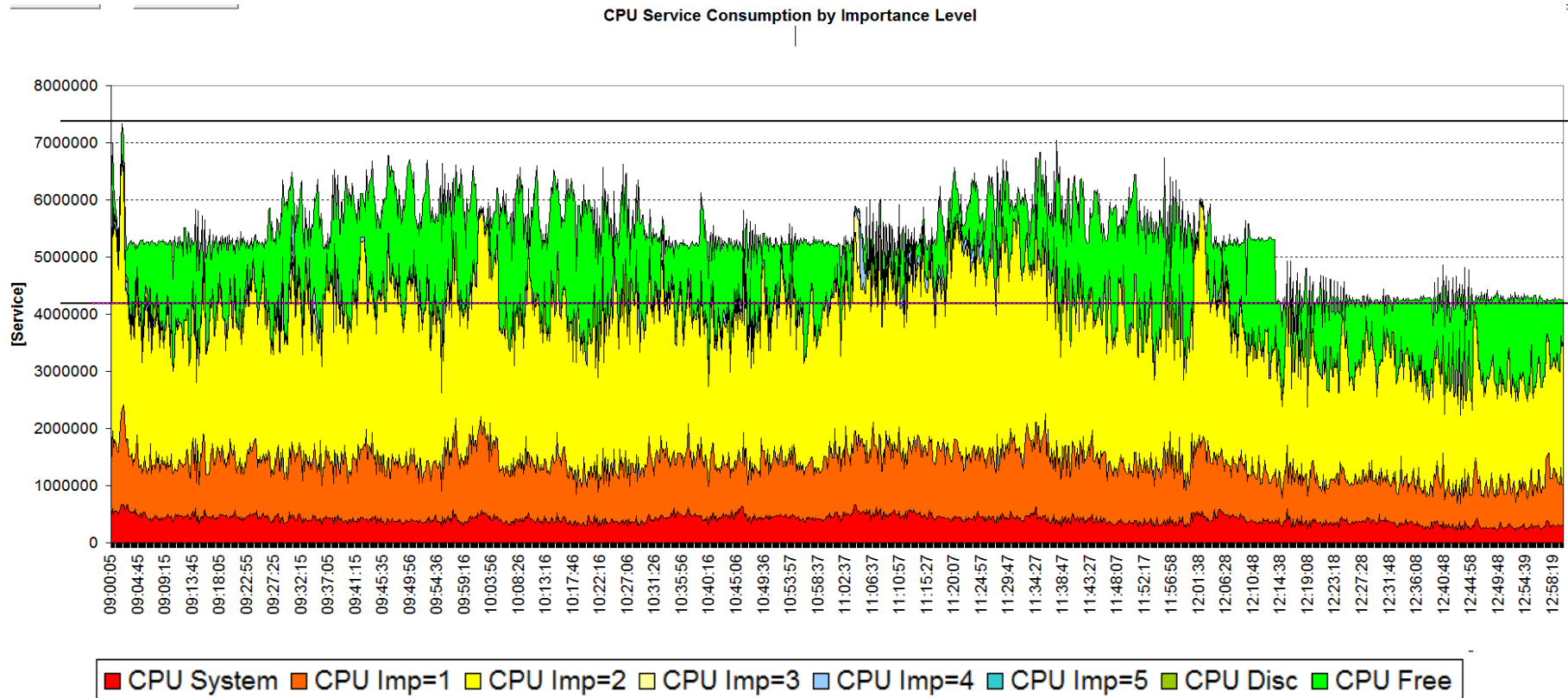
# Free LPAR Capacity - Example 1

zIIP Service Consumption by Importance Level



- While an LPAR is running below its weight entitlement and no capping is in effect the total consumed plus free capacity is usually pretty constant.

## Free LPAR Capacity - Example 2



- Capping, group capping, and influences by other LPARs can **heavily** and **frequently** change the total capacity available to an LPAR

## Free LPAR Capacity – some considerations


- *A single capacity value can hardly represent all the different preferences that an installation may have.*

Examples:

- Preferentially displace the lowest importance work
- Minimize/control crossover of zIIP/zAAP work to CPs
- Anticipation of capping before capping becomes active
- Equal distribution of used capacity
- Preferential use of guaranteed capacity vs. free CEC capacity
- Leave whitespace for anticipated additional workloads, e.g. batch
- Availability/anticipation of not yet activated temporary capacity (On/Off Capacity on Demand)
- Avoid using activated temporary capacity
- ...

**Blue: Controls are available in WLM, or routing services**

# Agenda

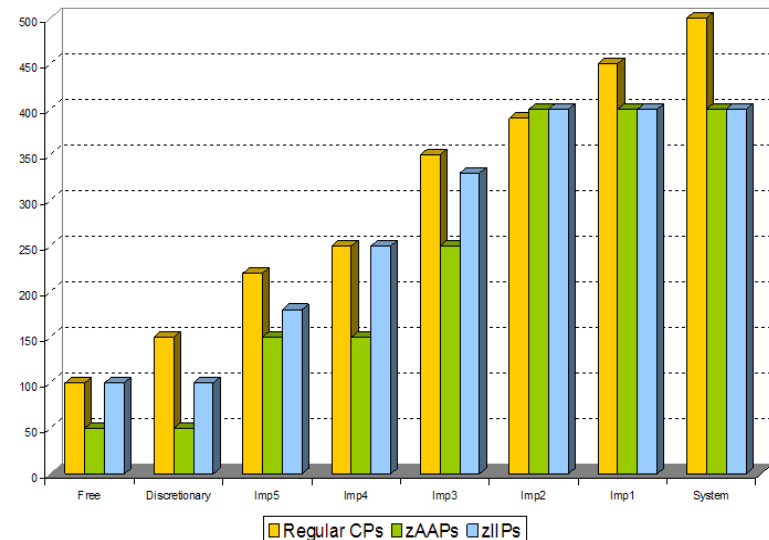
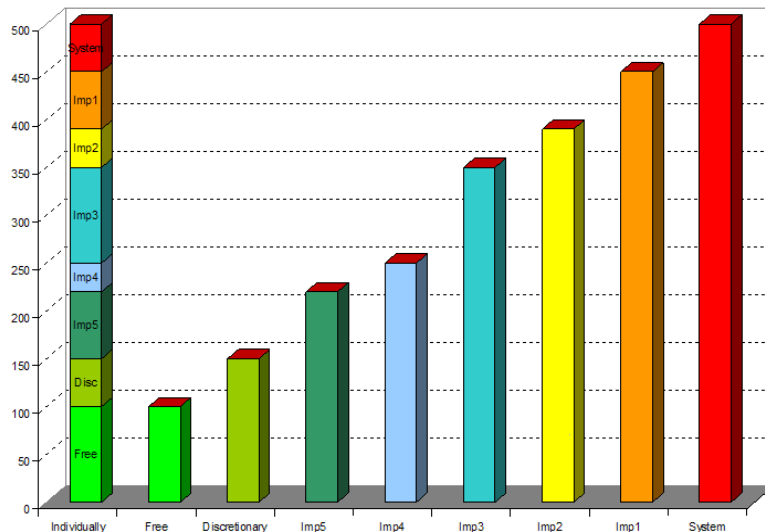
- Concepts
  - Importance levels
  - Displaceable capacity
  - Free capacity
-  ▪ WLM Sysplex Routing Services
  - IWMWSYSQ
  - IWMSRSRS
  - IWM4SRSC
  - Basic capacity-based weights and additional influencers
- Observations, best practices and optimization approaches

## WLM Sysplex Routing Services Overview

WLM Interface	Typical Use ( not exhaustive)	Purpose
IWMWSYSQ	Applications and sub-systems that want to consider free and displaceable capacity.	Obtain free & displaceable capacity of systems in Sysplex (1, 3, and 10 min rolling averages).
IWMSRSRS FUNCTION=SELECT (IWMSRSRG,DRS)	Sysplex Distributor BASEWLM	Obtain best suited registered servers to route work to. Only capacity considered.
IWMSRSRS FUNCTION=SPECIFIC (IWMSRSRG,DRS)	DDF	Obtain list of registered eligible servers and recommended weights. Considers capacity goal achievement (PI), queue time for enclaves, health.
IWMSRCRI	WebSphere	Similar to IWMSRSRS SPECIFIC but allow to concentrate work on application control regions.
IWM4SRSC	Sysplex Distributor SERVERWLM	Obtain recommendation for a specific server address space. No registration required. Capacity, server-specific capacity goal achievement (PI), abnormal termination rate, health is considered; optionally crossover cost and importance level weighting
IWM4HLTH	CICS Transaction Gateway, DB2/DDF, LDAP.	Provide health status for an address space. Value is considered by IWM4SRSC and IWMSRSRS FUNCTION=SPECIFIC

## Routing Services: IWMWSYSQ

- Provides displaceable capacity at each importance level
  - The system level contains the total system capacity, including SYSTEM work
  - Rolling average over 60, 180, and 600 sec.
- Data are returned for all processor types
- In addition: System shortages information, uniprocessor speed of a single processor, zAAP and zIIP normalization factors– required for subcapacity models
  - Use EXTENDED\_DATA=YES for comprehensive information



# Sample IWMWSYSQ Output

## - Only 180 sec CP and zIIP rolling average shown-

A	B	F	G	H	I	J	K	L	M	N	O	R	S	T	U	V	W	X	Y
LPAR	Time	CPU	CPI	G0_180	G1_180	G2_180	G3_180	G4_180	G5_180	GD_180	GF_180	I0_180	I1_180	I2_180	I3_180	I4_180	I5_180	ID_180	IF_180
SYS1	2016-03-31-11.18.	71428	3	1.8E+07	1.2E+07	1.2E+07	1.2E+07	1.2E+07	1.2E+07	1.1E+07	1.1E+07	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYS2	2016-03-31-11.18.	57971	20	7.3E+07	6.8E+07	5.9E+07	3.4E+07	3.3E+07	3.0E+07	2.7E+07	2.7E+07	3.5E+07	3.5E+07	2.9E+07	2.6E+07	2.6E+07	2.5E+07	2.5E+07	2.5E+07
SYS3	2016-03-31-11.18.	60836	17	9.4E+07	8.6E+07	6.5E+07	1.8E+07	1.8E+07	1.7E+07	1.4E+07	1.4E+07	5.0E+07	5.0E+07	4.2E+07	3.3E+07	3.3E+07	3.2E+07	3.2E+07	3.2E+07
SYS4	2016-03-31-11.18.	29850	4	9.0E+06	4.2E+06	4.1E+06	4.1E+06	4.1E+06	4.0E+06	3.8E+06	3.8E+06	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYS4	2016-03-31-11.18.	58823	19	9.8E+07	8.9E+07	6.8E+07	2.1E+07	1.4E+07	1.2E+07	1.8E+06	1.8E+06	2.6E+07	2.6E+07	2.0E+07	1.4E+07	1.4E+07	1.3E+07	1.3E+07	1.3E+07
SYS5	2016-03-31-11.18.	55555	22	1.1E+08	1.1E+08	9.0E+07	3.6E+07	3.3E+07	3.1E+07	2.8E+07	2.8E+07	4.5E+07	4.5E+07	3.5E+07	9.2E+06	9.1E+06	9.0E+06	9.0E+06	9.0E+06
SYS6	2016-03-31-11.18.	68085	10	2.2E+07	1.8E+07	1.7E+07	1.2E+07	9.6E+06	6.3E+06	3.5E+06	3.5E+06	3.5E+07	3.5E+07	3.3E+07	1.2E+07	1.2E+07	1.2E+07	1.1E+07	1.1E+07
SYS7	2016-03-31-11.18.	28933	6	7.6E+06	6.4E+06	6.1E+06	6.1E+06	6.1E+06	6.1E+06	6.1E+06	6.1E+06	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYS8	2016-03-31-11.18.	63241	13	5.2E+07	4.7E+07	4.0E+07	1.9E+07	1.6E+07	1.6E+07	1.5E+07	1.5E+07	1.5E+07	1.5E+07	1.1E+07	6.8E+06	6.8E+06	6.8E+06	6.6E+06	6.6E+06
SYS9	2016-03-31-11.18.	68085	8	3.6E+07	3.2E+07	2.6E+07	1.1E+07	6.7E+06	4.6E+06	2.2E+05	2.2E+05	1.9E+07	1.9E+07	1.4E+07	9.7E+06	9.7E+06	9.6E+06	8.7E+06	8.7E+06
SYSA	2016-03-31-11.18.	57971	20	1.1E+08	1.0E+08	8.7E+07	2.8E+07	2.6E+07	2.3E+07	2.0E+07	2.0E+07	4.1E+07	4.1E+07	3.4E+07	2.0E+07	2.0E+07	2.0E+07	2.0E+07	2.0E+07
SYSB	2016-03-31-11.18.	59701	12	6.5E+06	5.5E+06	5.3E+06	5.2E+06	5.1E+06	4.9E+06	4.7E+06	4.7E+06	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYSC	2016-03-31-11.18.	66945	6	8.7E+06	7.8E+06	7.8E+06	7.8E+06	7.8E+06	7.8E+06	7.8E+06	7.8E+06	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYSD	2016-03-31-11.18.	69868	10	1.1E+07	8.4E+06	7.5E+06	4.3E+06	4.2E+06	4.1E+06	2.6E+04	2.6E+04	2.7E+07	2.7E+07	2.6E+07	1.1E+07	1.1E+07	1.1E+07	1.0E+07	1.0E+07
SYSE	2016-03-31-11.18.	69868	10	1.4E+07	1.2E+07	1.1E+07	8.8E+06	8.0E+06	8.0E+06	8.0E+06	8.0E+06	4.9E+07	4.9E+07	4.9E+07	3.1E+07	3.1E+07	3.1E+07	3.1E+07	3.1E+07
SYSF	2016-03-31-11.18.	64777	8	7.5E+06	5.6E+06	5.0E+06	5.0E+06	4.8E+06	4.2E+06	7.8E+05	7.8E+05	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYT1	2016-03-31-11.18.	59701	12	1.2E+07	1.1E+07	1.1E+07	1.1E+07	1.1E+07	1.1E+07	9.7E+06	9.7E+06	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYT2	2016-03-31-11.18.	65843	7	4.7E+06	3.6E+06	3.6E+06	3.5E+06	3.4E+06	3.4E+06	3.3E+06	3.3E+06	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYT4	2016-03-31-11.18.	28933	6	1.9E+06	9.0E+05	8.3E+05	7.1E+05	6.7E+05	6.7E+05	5.4E+05	5.4E+05	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
SYS5	2016-03-31-11.18.	28520	7	3.4E+06	2.3E+06	2.3E+06	2.1E+06	2.1E+06	2.1E+06	2.0E+06	2.0E+06	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00

- In this spreadsheet view blue (CP) and green (zIIP) bars indicate the amount of displaceable capacity at each level
  - Each column scaled separately, representing the ratio at a given importance
- G0 is basically the LPAR capacity, GF the free capacity.
  - Depending on the workload consumption within the LPARs, ratios change between importance levels.

# WLM Routing Weights Computation

## Overview: *Steps Involved*

1. Compute system-level weights based on capacity view
  - Compute weights for each processor type and combined weight
  - Frequently scaled to 64
  - Optionally, apply adjustments for crossover cost, and importance level weighting
2. When multiple servers run on a system divide the system weight by #servers to derive a server's weight
3. Only for IWMSRSRS SPECIFIC and IWM4SRSC, modify weights based on
  - Performance index
  - Queue time ratio
  - Health indicator
  - For IWM4SRSC, consider the abnormal termination rate



## IWMSRSRS vs. IWM4SRSC Capacity calculations

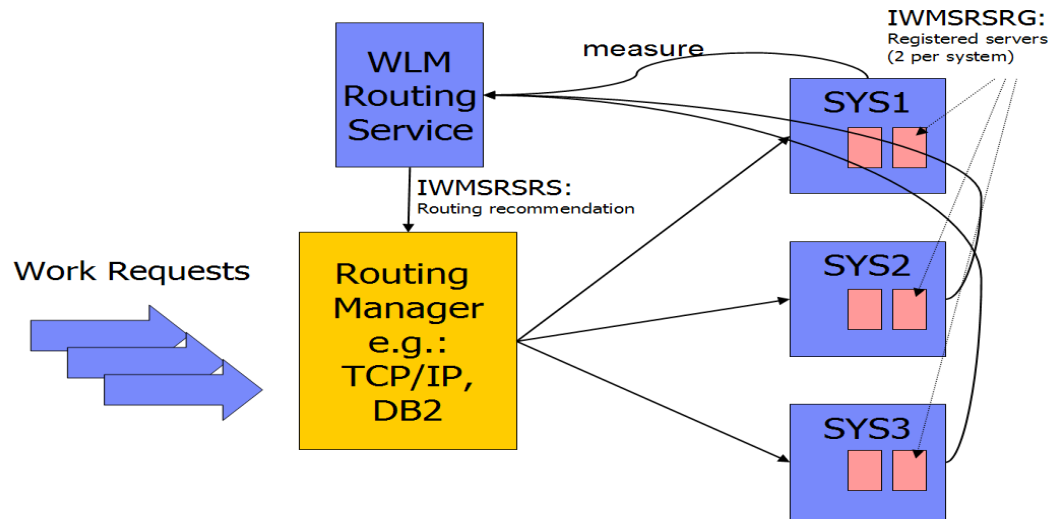
### WLM service IWMSRSRS (SD BASEWLM or DDF)

- Locate the importance level –searching bottom-up- where at least 5% of free/ displaceable capacity is available on one system
- Possible disadvantages
  - Importance of work to be routed is not considered
  - May result in oscillations that would usually smooth out over time, though
  - Only SD BASEWLM: recommendations are for the system; no server specific insights
- Advantage
  - Considers the low important work because it is a bottom up approach; i.e., will be “pushed aside” even by low priority work
  - Tends to result in more equal distribution of free capacity

### WLM service IWM4SRSC (SD SERVERWLM)

- Calculates the weight based on the *displaceable capacity at the importance level that the work will run on the systems.*
- Possible disadvantage
  - Less important work is not distinguished from free capacity (But → Importance Level Weighting)
- Advantages
  - Considers the importance of the work. Tends to push aside lower importance work.
  - Avoids the oscillation of routing recommendations
  - Tends to optimize workload performance

# Sysplex Routing with IWMSRSRS: Bottom-Up Weight Calculation



## Algorithm

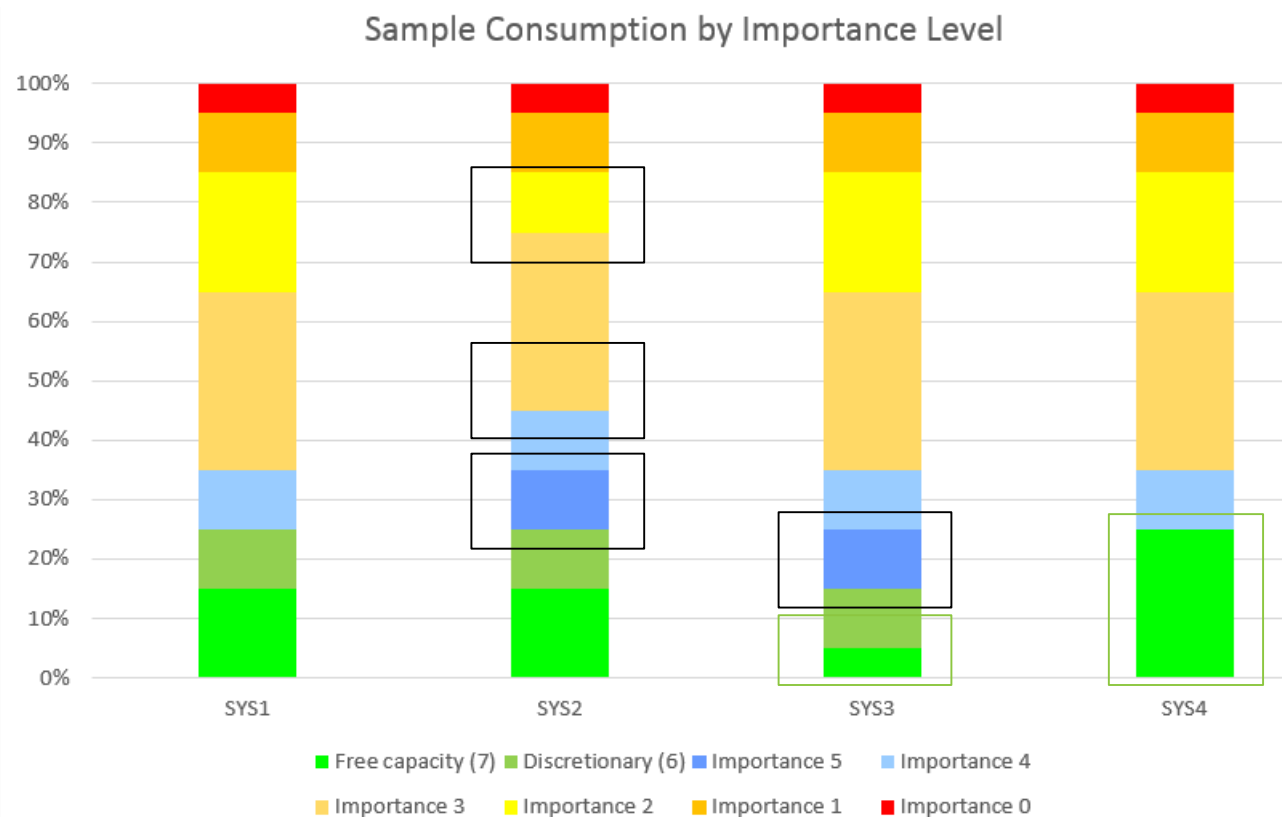
1. Select the importance level that provides at least **5% of cumulative capacity on at least one system**
2. Calculate system weight on each system
3. Calculate server weight:

$$\text{System Weight} = \frac{\text{SUs}_{\text{at selected level [this system]}}}{\sum_{\text{I for all systems}} \text{SUs}_{\text{at selected level [i]}}} \cdot 64$$

$$\text{Server Weight} = \frac{\text{System Weight}}{\text{\# of servers on system}}$$

## Example: System weights for IWMSRSRS and IWM4SRSC

- Assumptions:
  - Four systems identically configured → same total LPAR capacity
  - Similar consumptions patterns
    - But different consumption at importance level 2, 4 and discretionary (6)
- Questions:
  - What recommendations will be given by the different routing services?
  - What may be the consequences?



Rectangles highlight differences in workload, and resulting free capacity

## Example: System weights for IWMSRSRS (“SD BASEWLM” & DDF)

System	SYS1		SYS2		SYS3		SYS4		Total avail. SU at Importance	Total displace- able capacity
	kSU	Cum. kSU	kSU	Cum. kSU	kSU	Cum. kSU	kSU	Cum. kSU		
LPAR capacity	100		100		100		100			
Importance 0	5	100	5	100	5	100	5	100	20	400
Importance 1	10	95	10	95	10	95	10	95	40	380
Importance 2	20	85	10	85	20	85	20	85	70	340
Importance 3	30	65	30	75	30	65	30	65	120	270
Importance 4	10	35	10	45	10	35	10	35	40	150
Importance 5	0	25	10	35	10	25	0	25	20	110
Discretionary (6)	10	25	10	25	10	15	0	25	30	90
Free capacity (7)	15	15	15	15	5	5	25	25	60	60
IWMSRSRS system weight recommendation	16		16		5		27			

kSU = 1000 Service Units = Technical measure of CPU capacity

- Selected importance level: 7 (more than 5% of 100 kSU available)
- SYS1: Weight =  $15 * 64 / 60 = 16$
- SYS2: Weight =  $15 * 64 / 60 = 16$
- SYS3: Weight =  $5 * 64 / 60 = 5$
- SYS4: Weight =  $25 * 64 / 60 = 27$

## Example: System weights for IWM4SRSC (SD “SERVERWLM”)

System	SYS1		SYS2		SYS3		SYS4		Total avail. SU at Importance	Max. displace- able capacity
	kSU	Cum. kSU	kSU	Cum. kSU	kSU	Cum. kSU	kSU	Cum. kSU		
LPAR capacity	100		100		100		100			
Importance 0	5	95	5	95	5	95	5	95	20	95
Importance 1	10	85	10	85	10	85	10	85	40	85
Importance 2	20	65	10	75	20	65	20	65	70	75
Importance 3	30	35	30	45	30	35	30	35	120	45
Importance 4	10	25	10	35	10	25	10	25	40	35
Importance 5	0	25	10	25	10	15	0	25	20	25
Discretionary (6)	10	15	10	15	10	5	0	25	30	25
Free capacity (7)	15	0	15	0	5	0	25	0	60	0
IWM4SRSC system weight recommendation	50		64		50		50			

- Selected importance level: 3 (based on routed work)
- SYS1: Weight =  $35 * 64 / 45 = 50$
- SYS2: Weight =  $45 * 64 / 45 = 64$
- SYS3: Weight =  $35 * 64 / 45 = 50$
- SYS4: Weight =  $35 * 64 / 45 = 50$

## WLM Routing: Importance level weighting

- Importance Level Weighting is available with service IWM4SRSC (SD SERVERWLM)
- Addresses the concern that work of lower importance than the selected importance level is treated like free capacity
  - Allow to differentiate between lower importance levels. Free capacity and very low importance work can be preferentially displaced
- Four weighting levels exist:
  - IL0: Default uses “Constant” - no weighting of the lower importance levels
  - IL1: Square Root (mildly – recommended initial setting),
  - IL2: Linear
  - IL3: Quadratic (heavy) weighting
- In the examples, you can observe that the biggest effect is for system J3 on which much more work runs at importance level 2

- If the Performance Index (PI) >1 the weight will be divided by MAX(PI, 5.0 )
  - Weights systems with over-achieving work will not be increased
- IEAOPT RTPIFACTOR allows to scale back the effect of the PI

System	Avail Capacity	Orig. Server weight	PI	WLM weight
SYS1	110	18	1.3	14
SYS2	100	16	0.8	16
SYS3	95	15	1.0	15
SYS4	95	15	2.0	8
Total		64		53

## Health indicator effect on routing weight

- A health indicator may be set per for a –server- address space
- Health=100 is default and remains in effect until a different value is set via IWM4HLTH
  - Up to z/OS V2.1: Each IWM4HLTH invocation replaces previous health indicator values
  - z/OS V2.2 and above: Multiple providers can provide their view of the health. The aggregated health value (minimum of all values provided) is used for the weight calculation.
- If the health indicator of a server is <100 its capability is reduced
- The server weight will be reduced by applying a factor of health/100
- IWM4SRSC also considers the ratio abnormal:normal transaction completions, as reported by the subsystem.

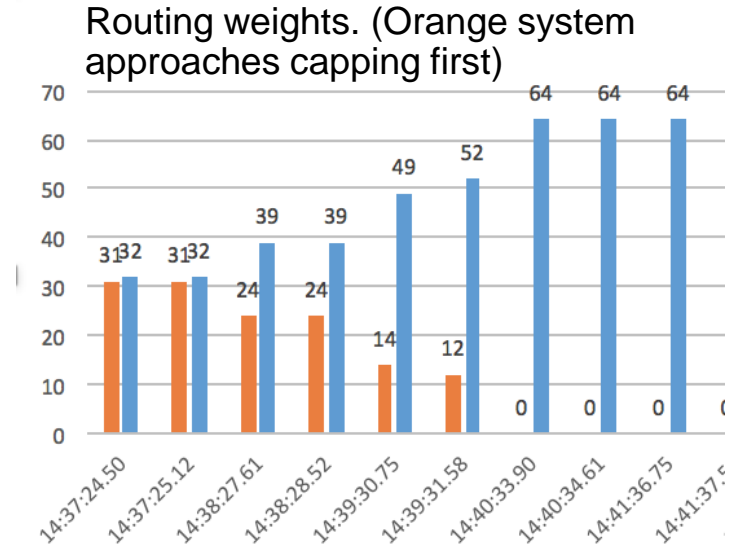


## Time-to-cap aware routing

- As of z/OS V2.3 routing recommendations can optionally reflect capping before capping actually becomes effective.

Possible advantages:

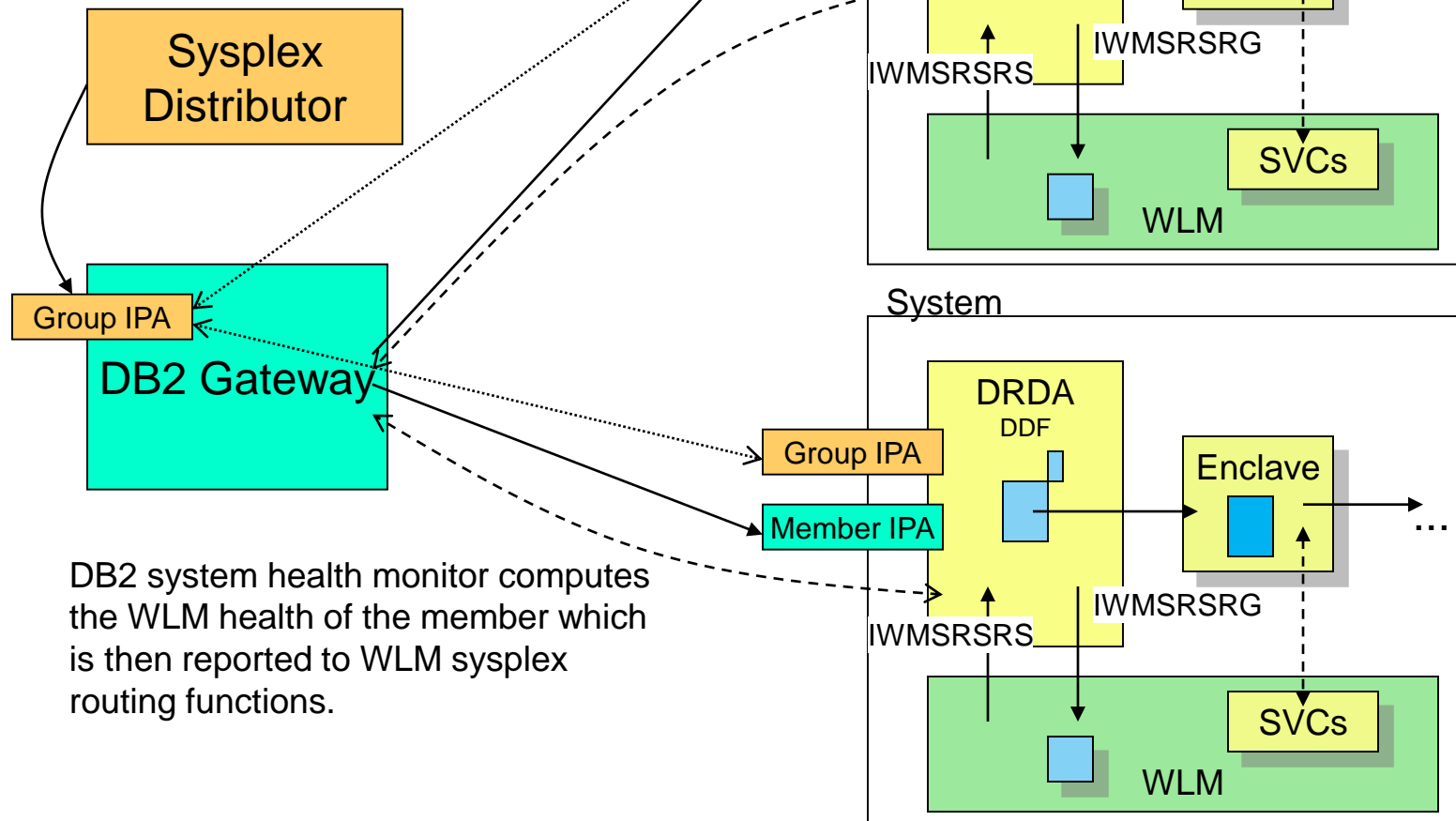
- More balanced use of capacity within the Sysplex
  - Less queueing of work when capping hits
- Can be enabled per system in the IEAOPTxx member by specifying a non-zero value for RtCapLeadTime.
    - Values between 3 and 20 are likely best suited.

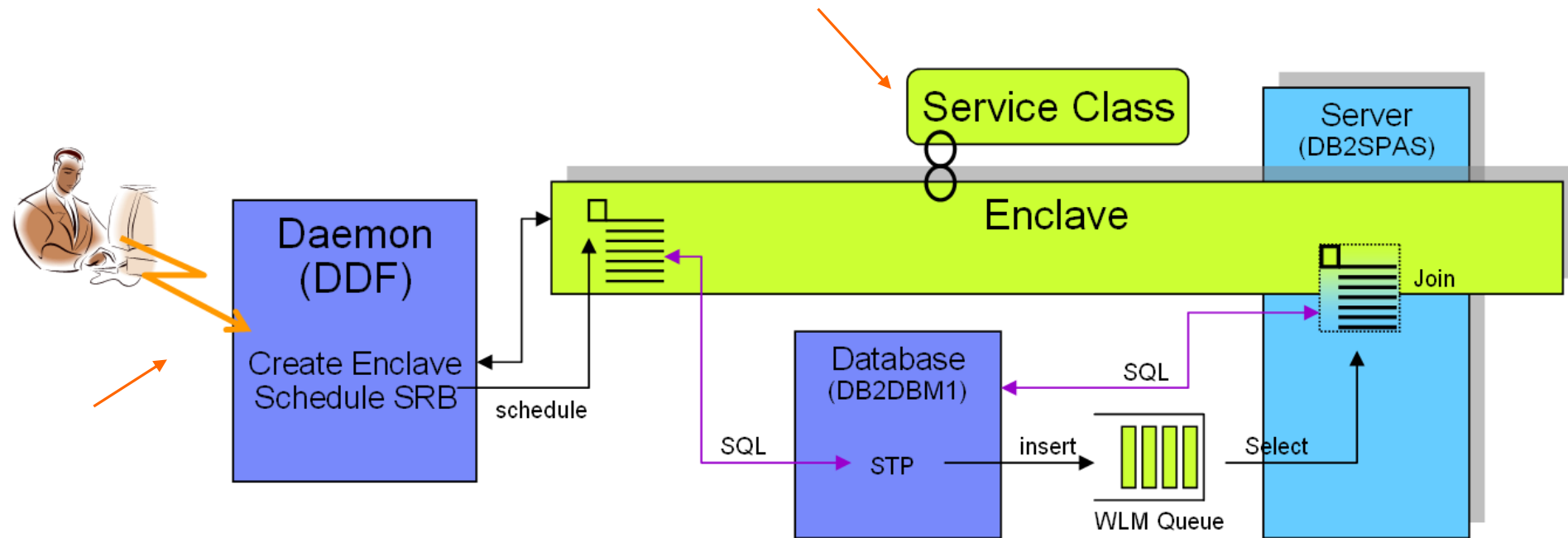


IEAOPTxx RTCAPLEADTIME=n n	
<u>0</u>	Default behavior is to not consider soft capping in advance. Capping that is already in effect is always reflected.
[1-60]	Specifies the time in minutes, how long in advance an upcoming soft capping should influence WLM's sysplex routing recommendations. When the estimated time to capping is less than n minutes WLM reflects the upcoming soft capping in it's routing recommendations.

## Background: Routing Services: DB2

DDF address spaces register as routing servers to WLM. DDF address spaces also periodically retrieve the routing list and ship it to the gateway which routes the requests.



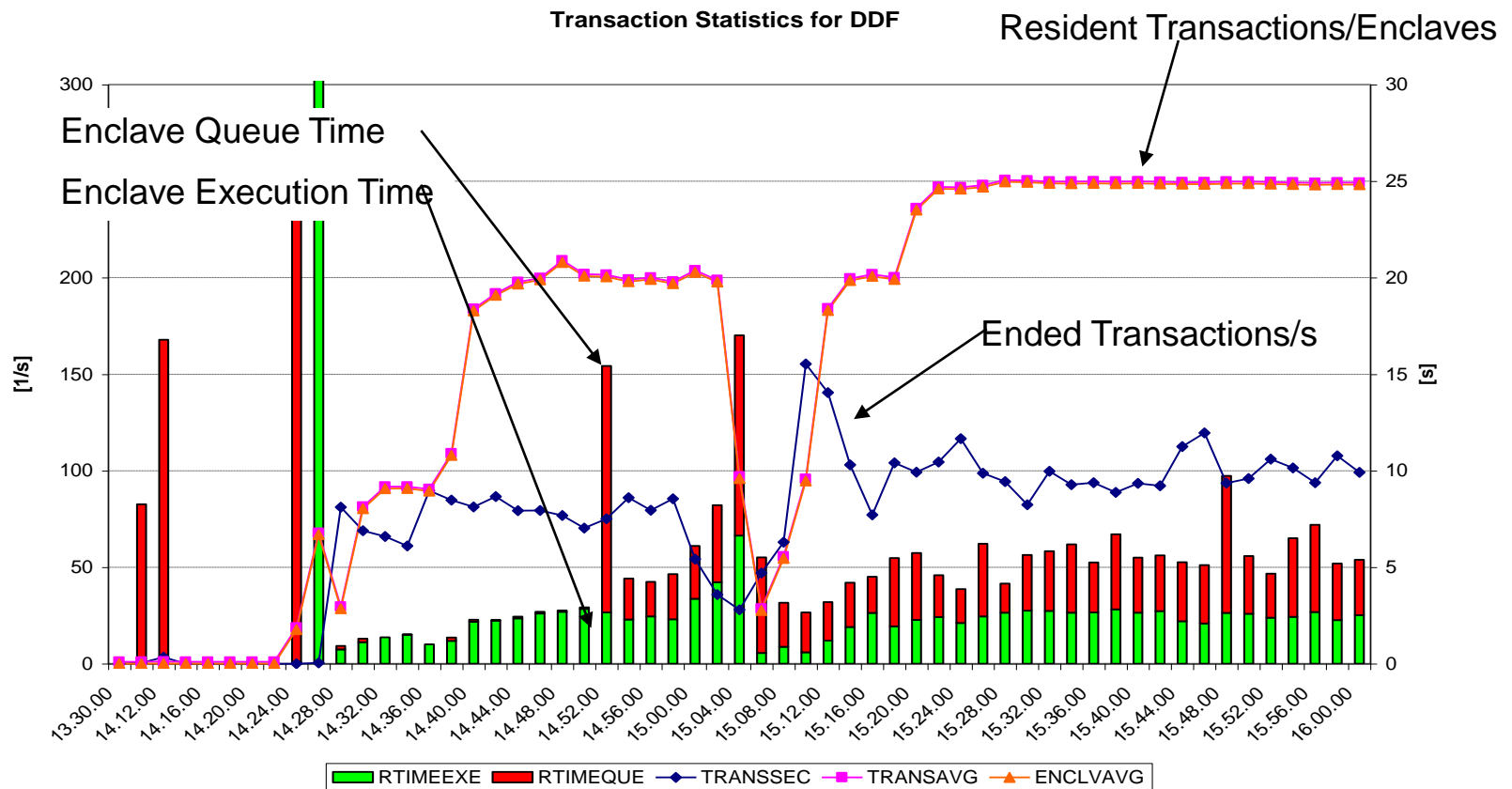


IWMSRSRS FUNCTION=SPECIFIC does also consider

- The performance index that indicates the achievement of the WLM defined goals of the server, that is its related work.
- If the server owns independent enclaves those also take the delays into account that the work is subject to, due to the queue times of the owned enclaves.
- Health factor reported for a server

## Sysplex Routing for DB2: Example Queue Time Ratio

- Servers with a better enclave queue time : execution time ratio will be favored
  - Server weight reduced by factor  $\text{execution time} / (\text{execution time} + \text{queue time})$
- Only effective if DB2 is configured with "DDF Threads" INACTIVE



# Agenda

---

- Concepts
  - Importance levels
  - Displaceable capacity
  - Free capacity
- WLM Sysplex Routing Services
  - IWMWSYSQ
  - IWMSRSRS
  - IWM4SRSC
  - Basic capacity-based weights and additional influencers



- Observations, best practices and optimization approaches

- Actual workload distribution may deviate from anticipated or desired distribution
  - When “desired” – why?
- Understanding and optimizing workload routing may require skills from multiple domains:
  - Applications
  - Subsystems involved
  - Routing product & configuration
    - Routing provide usually commands to understand WLM-provided weights and overrides
    - First step is to understand raw WLM weights
    - Most routing services parameters are specified here
  - LPAR configuration & WLM

## Drill-down into balancing issues

- What routing product and service is being used?
- Use routing component commands to understand WLM recommendations vs. routed work
  - A good approach is to issue the commands every minute or few minutes and record the output.
    - Besides WLM weights, also the health is reported
  - Is already the WLM recommendation “unexpected”, or are the weights reasonable but the workload distribution is different?
- If WLM weight related, understand impacts due to
  - Capacity
  - Performance Index
  - Health

Use CPU activity report and Workload activity reports to understand LPAR/CEC configuration, load and performance index

  - RMF Mon III data can provide better granularity

# TCPIP Sysplex Distributor analysis

## NETSTAT -O



```
$ netstat -O -  
P15150
```

```
MVS TCP/IP NETSTAT CS
```

```
V1R12      TCPIP Name:  
TCPIP      10:31:18
```

```
Dynamic VIPA Destination Port Table  
for TCP/IP stacks:
```

```
Dest:      ....15150
```

```
DestXCF:    ...
```

```
TotalConn: 0000059767 Rdy: 001
```

```
WLM: 12  TSR: 100
```

```
DistMethod: ServerWLM
```

- The WLM weight in this summary display is derived by the weight value returned by IWM4SRSC (ServerWLM)
  - However, it has been post processed by Sysplex Distributor
    - Potentially reduced based on a number of health factors and
    - Normalized (divided by 4 to yield a value between 0-16 vs 0-64).
  - This value is what SD will use for load balancing and can be compared to the values of the other targets
- TSR (target server responsiveness) the SD view of responsiveness of target servers in accepting new connections. The TSR values are used to modify the weight used to favor servers that are more successfully accepting new connection requests. A value of 100 indicates full responsiveness and zero indicates no responsiveness.



# TCPIP Sysplex Distributor analysis

## NETSTAT VIPADCFG DETAIL



### VIPA Distribute:

IP Address	Port	XCF Address	SysPt	TimAff	Flg
-----	----	-----	-----	-----	-----
201.2.10.11	n/a	ALL	Yes	200	R
DistMethod: Roundrobin					
OptLoc: No					
201.2.10.13	243	ALL	No	No	0
DistMethod: BaseWLM					
OptLoc: 1					
ProcType:					
CP: 60 zAAP: 00 zIIP: 40					
201.2.10.14	243	ALL	No	No	1
DistMethod: ServerWLM					
OptLoc: No					
ProcXCost:					
zAAP: 003 zIIP: 001					
ILWeighting: 1					

# DB2 DDF analysis

## DDF DISPLAY Command



- -DIS DDF [DETAIL] returns WLM weight information
  - The following server list entry information is displayed for each DDF location that registered to WLM as part of the data sharing group:
  - DSNL100I LOCATION SERVER LIST: DSNL101I WT IPADDR IPADDR DSNL102I  
*weight ipv4-address ipv6-address*

```
-DISPLAY DDF DETAIL
```

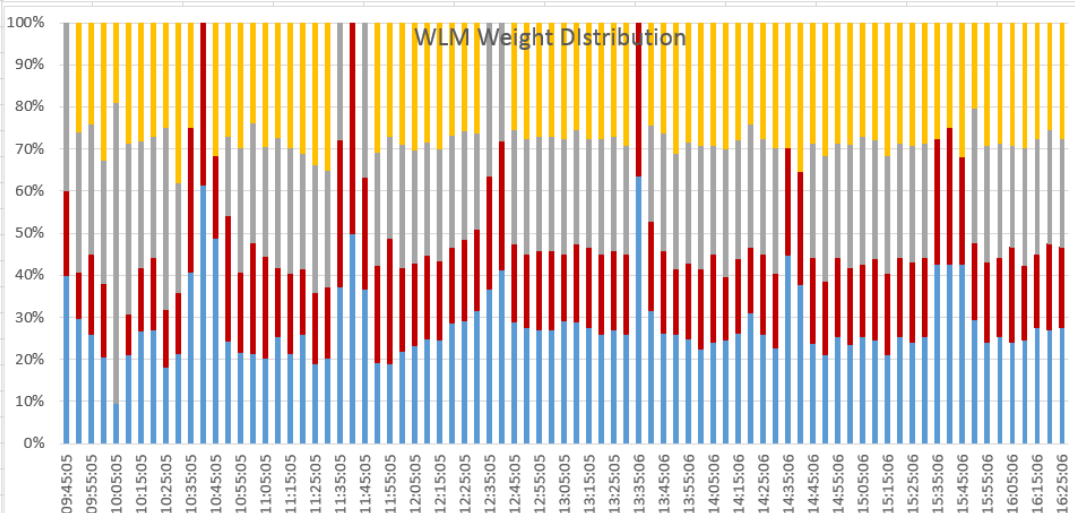
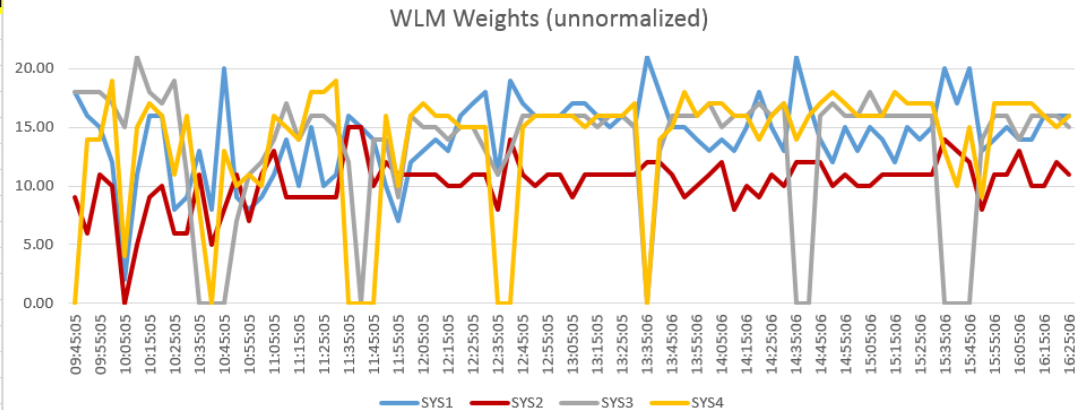
With the DETAIL option, the following additional information is included in the output:

```
DSNL090I DT=      A CONDBAT=  64 MDBAT=   64
DSNL092I ADBAT=   1 QUEDBAT=   0 INADBAT=   0 CONQUED=      0
DSNL093I DSCDBAT= 0 INACONN=   0
DSNL100I      LOCATION SERVER LIST:
DSNL101I      WT IPADDR      IPADDR
DSNL102I      64 ::9.110.115.111  2002:91E:610:1::111
DSNL102I      ::9.110.115.112  2002:91E:610:1::112
DSNL099I DSNLTDDF DISPLAY DDF REPORT COMPLETE
```

# Sample DDF analysis

- The output of the commands can easily be tabled and analyzed. Compare with actual workload distribution to verify

Time	9.132.1.31	9.132.1.32	9.132.1.33	9.132.1.34
09:45:05	18.0	9.0	18.0	
09:50:05	16.0	6.0	18.0	14.0
09:55:05	15.0	11.0	18.0	14.0
10:00:05	12.0	10.0	17.0	19.0
10:05:05	2.0		15.0	4.0
10:10:05	11.0	5.0	21.0	15.0
10:15:05	16.0	9.0	18.0	17.0
10:20:05	16.0	10.0	17.0	16.0
10:25:05	8.0	6.0	19.0	11.0
10:30:05	9.0	6.0	11.0	16.0
10:35:05	13.0	11.0		8.0
10:40:05	8.0	5.0		
10:45:05	20.0	8.0		13.0
10:50:05	9.0	11.0	7.0	10.0
10:55:05	8.0	7.0	11.0	11.0
11:00:05	9.0	11.0	12.0	10.0
11:05:05	11.0	13.0	14.0	16.0
11:10:05	14.0	9.0	17.0	15.0
11:15:05	10.0	9.0	14.0	14.0
11:20:05	15.0	9.0	16.0	18.0
11:25:05	10.0	9.0	16.0	18.0
11:30:05	11.0	9.0	15.0	19.0
11:35:05	16.0	15.0	12.0	
11:40:05	15.0	15.0		
11:45:05	14.0	10.0	14.0	
11:50:05	10.0	12.0	14.0	16.0
11:55:05	7.0	11.0	9.0	10.0
12:00:05	12.0	11.0	16.0	16.0
12:05:05	13.0	11.0	15.0	17.0
12:10:05	14.0	11.0	15.0	16.0
12:15:05	13.0	10.0	14.0	16.0
12:20:05	16.0	10.0	15.0	15.0
12:25:05	17.0	11.0	15.0	15.0
12:30:05	18.0	11.0	13.0	15.0
12:35:05	11.0	8.0	11.0	
12:40:05	19.0	14.0	13.0	



- The DB2 health value can be obtained via the following messages:
  - **DISPLAY THREAD(\*) TYPE(SYSTEM)** command will issue message DSNV507I (ACTIVE MONITOR...)
  - **DISPLAY DDF DETAIL** command will issue DSNL094I when the subsystem is a member of a data sharing group.

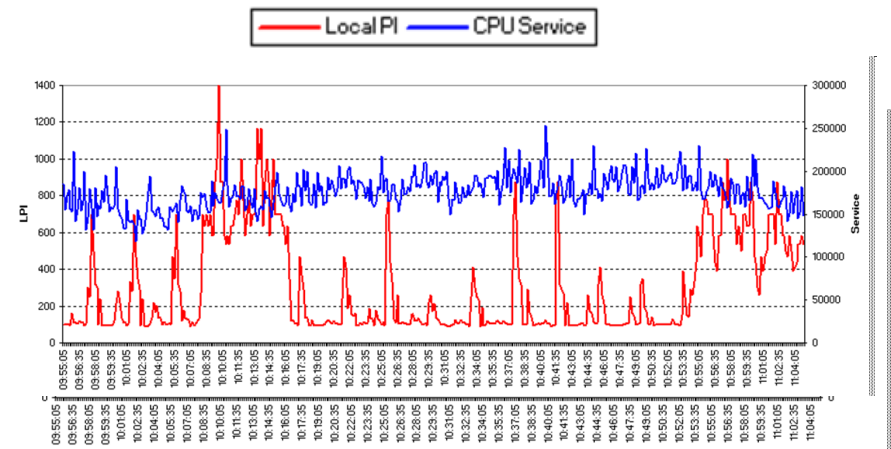
- On z/OS V2.2 and above also via RunTime Diagnostic: F HZR,ANALYZE

```
-----
EVENT 17: HIGH - SERVERHEALTH - SYSTEM: SYS1      2016/04/19 - 08:00:30
JOB NAME: DB1XDIST  ASID: 01CC  CURRENT HEALTH VALUE: 75
CURRENT LOWEST HEALTH VALUES:
      SUBSYSTEM  HEALTH              REPORTED
SUBSYSTEM NAME    SETTING          REASON  DATE AND TIME
DB1TDIST          75                2016/04/19 06:01:04
ERROR: ADDRESS SPACE SERVER CURRENT HEALTH VALUE LESS THAN 100.
ERROR: THIS VALUE MAY IMPACT YOUR SYSTEM OR SYSPLEX TRANSACTION
ERROR: PROCESSING.
ACTION: USE YOUR SOFTWARE MONITORS TO INVESTIGATE THE ASID AND TO
ACTION: DETERMINE THE IMPACT OF THE HEALTH OF THE ADDRESS SPACE TO
ACTION: OVERALL TRANSACTION PROCESSING.
-----
```

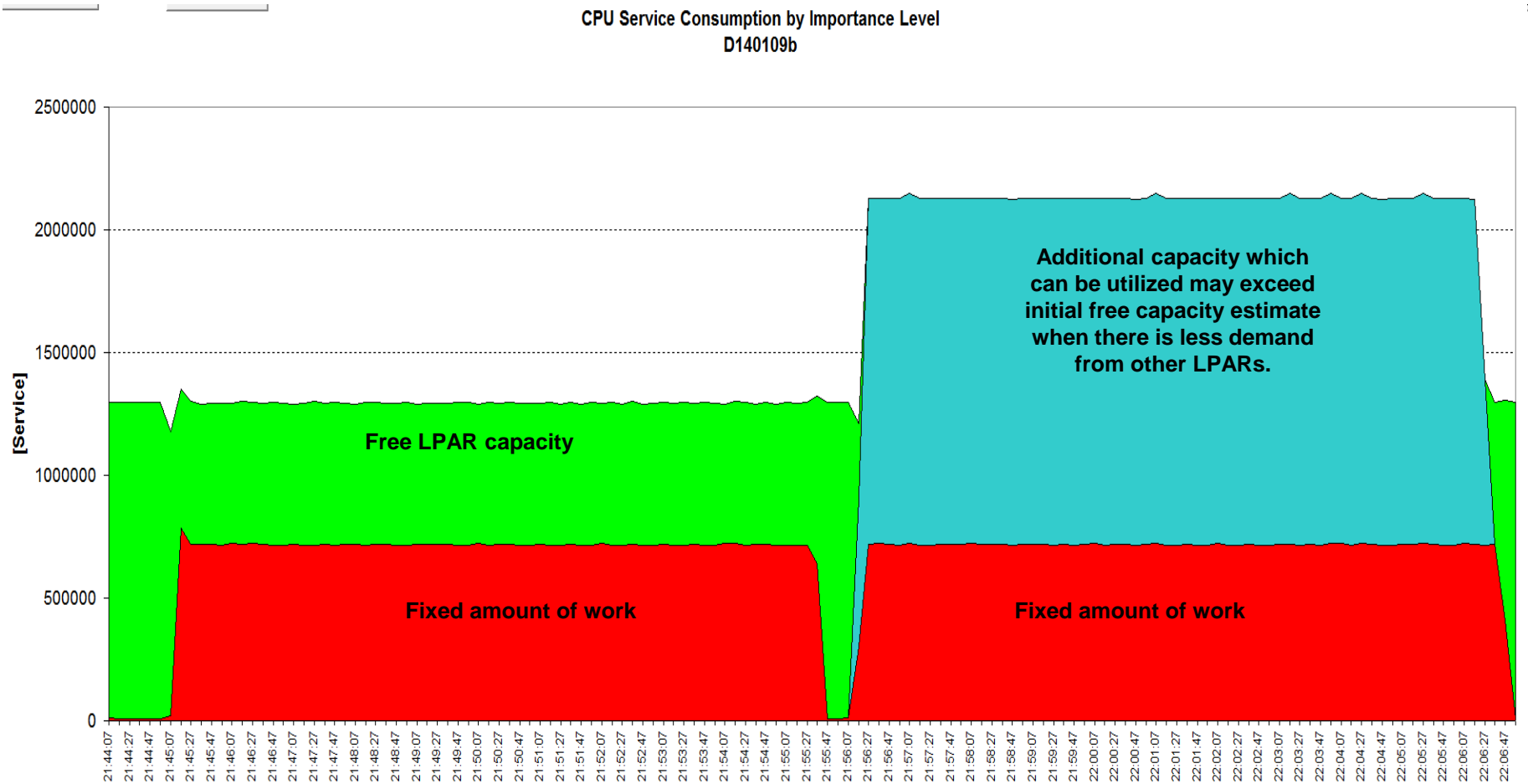
- DB2 health value depends on number of connections and storage utilization. For example,
  - When #Connections > 80% of CONDBAT: Health divided by 2
  - When #Connections > 90% of CONDBAT: Health divided by 4
  - **dis thd(\*) type(system)** can provide addition detail in “ACTIVE MONITOR” line
  - Additional information is available in this APAR:  
<http://www.ibm.com/support/docview.wss?uid=swg1PM43293>

## Performance Index (PI) effect on routing weight

- Heavily fluctuating high PI values can influence routing recommendations
  - Sometimes more than desired from a routing perspective
- Or some systems may show systematically higher PIs than other systems for unrelated reasons
- In such cases it can be beneficial to scale back the impact of the PI via the IEAOPT RTPIFACTOR control.
- When RTPIFACTOR=0, the server weight is independent from the server or PI
- When server PI >1 and
  - RTPIFACTOR=100 , the server weight is divided by the server PI
  - $0 < \text{RTPIFACTOR} < 100$  it results in a proportional influence of the server PI on the server weight.



## Example: Initial Free LPAR capacity may be under-estimated



## Observation: Connections vs. transaction routing

- Long living connections are... long living
  - Established at one point in time due to the load distribution at that time but not redistributed until connections are broken up and re-established
- Distributed DB2 work can exhibit “affinity” to a certain member caused by application behavior.
  - For example, Open WITH HOLD cursors, existing, declared global temporary tables which have not been dropped prior to commit, keep dynamic packages...
- The number of transactions routed to some systems may be *not proportional* to the number of connections that were established
  - For example, MQ channels.

- Usually not a problem at all - unless a specific distribution is warranted
- Asymmetric configuration may result in biased weights
  - E.g. different weights, different CEC configurations
    - Consider zIIP, zAAP pools, too, when relevant
  - Depending on subsystems the routed transactions could deviate more
- Consider
  - SERVERWLM - if PI is a good indicator for overload
  - IL Weighting
    - IL weighting=1 is usually a good starting point
  - Round-robin or another, non-WLM based distribution method



## Sysplex Distributor and DB2 DDF - More Information -

- Gus Kassimis:  
Sysplex Networking Technologies and Considerations,  
SHARE in Pittsburgh, 2014, Session: 15507
  
- DB2 9 for z/OS: Distributed Functions  
<http://www.redbooks.ibm.com/abstracts/sg246952.html?Open>
  
- Jim Pickel:  
DB2 9 for z/OS Data Sharing: Distributed  
Load Balancing and Fault Tolerant Configuration  
<http://www.redbooks.ibm.com/abstracts/redp4449.html>

## z/OS Workload Management - More Information -

- z/OS WLM Homepage:  
<http://www.ibm.com/systems/z/os/zos/features/wlm/>
- z/OS MVS documentation
  - z/OS MVS Planning: Workload Management:  
<http://publibz.boulder.ibm.com/epubs/pdf/iea2w1c0.pdf>
  - z/OS MVS Programming: Workload Management Services:  
<http://publibz.boulder.ibm.com/epubs/pdf/iea2w2c0.pdf>
- *IBM Redbooks publications:*
  - System Programmer's Guide to: Workload Manager:  
<http://publib-b.boulder.ibm.com/abstracts/sg246472.html?Open>
  - ABCs of z/OS System Programming Volume 12  
<http://publib-b.boulder.ibm.com/abstracts/sg247621.html?Open>

### Workload Manager

Welcome to WLM/SRM



Overview

What's New

FAQs

Further Information

## What is a DDF Transactions?

- ACTIVE MODE threads are treated as a single enclave from the time they are created until the time they are terminated. This means that the entire life of the database access thread is reported regardless of whether SQL work is actually being processed.
- INACTIVE MODE threads are treated differently. If the thread is always active, the duration of the thread is the duration of the enclave. When the thread is pooled, such as during think time, it is not using an enclave. In this case, inactive periods are not reported.

