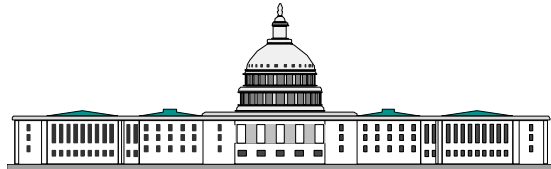


WSC Experiences with the zIIP Processor

Walt Caprice
Washington Systems Center
IBM
August 28, 2006



Trademarks

AIX*	IBM eServer	z/VM*
CICS*	IBM logo*	zSeries*
DB2*	IMS	
DB2 Connect	On Demand Business logo	
DB2 Universal Database	Parallel Sysplex*	
DRDA*	System z	
FICON*	System z9	
GDPS*	WebSphere*	
HiperSockets	z/Architecture	
IBM*	z/OS*	

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

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

Linux is a registered trademark of Linus 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.

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.

Disclaimer

The information in this document has not been submitted to any formal IBM test and is distributed on an "as is" basis without any warranty expressed or implied. Use of this information or the implementation of any of these techniques is a user responsibility and depends on the user's ability to evaluate and integrate them into the user's operational environment. While each item may have been reviewed for accuracy in a specific situation there is no guarantee the same or similar results may be achieved elsewhere. Users attempting to adapt these techniques to their environments do so at their own risk.

Agenda

- A zIIP overview
- What are the tuning options
- RMF updated reports
- Measurements
 - Run objectives
 - Results
- Summary
- Questions

A zIIP Overview

- New specialty engine for the System z9 mainframe designed to help:
 - **Customers integrate data across the enterprise**
 - **Improve resource optimization and lower the cost of ownership for eligible data serving workloads**
- z/OS manages and directs work between the general purpose processor and the zIIP
 - **No changes anticipated to DB2 for z/OS V8 applications**
 - **Number of zIIPs per z9 not to exceed number of standard processors**
- DB2 for z/OS V8 will be first exploiter of the zIIP with:
 - **System z9**
 - **z/OS 1.6 or later**
 - **DB2 for z/OS V8**
 - **Software GA 6/30/06**

DB2 Eligible Workloads

- ERP, CRM and multi tier application serving*
 - For applications, running on z/OS, UNIX, Linux, Intel, or Linux on System z, that access DB2 for z/OS V8 on a System z9, via DRDA over a TCP/IP connection DB2 gives z/OS the necessary information to have portions of these SQL requests directed to the zIIP
- Data warehousing applications*
 - Requests that utilize DB2 for z/OS V8 star schema parallel queries may have portions of these SQL requests directed to the zIIP when DB2 gives z/OS the necessary information.
 - This function has been enhanced to include all parallel queries
- Some DB2 for z/OS V8 utilities*
 - A portion of DB2 utility functions used to maintain index maintenance structures (example LOAD, REORG, and REBUILD INDEX) typically run during batch, can be redirected to zIIP.

* The zIIP is designed so a program can work with z/OS to have a portion of it's Service Request Block (SRB) enclave work directed to the zIIP. The above types of DB2 V8 work are those executing in SRB enclaves, portions of which can be sent to the zIIP.

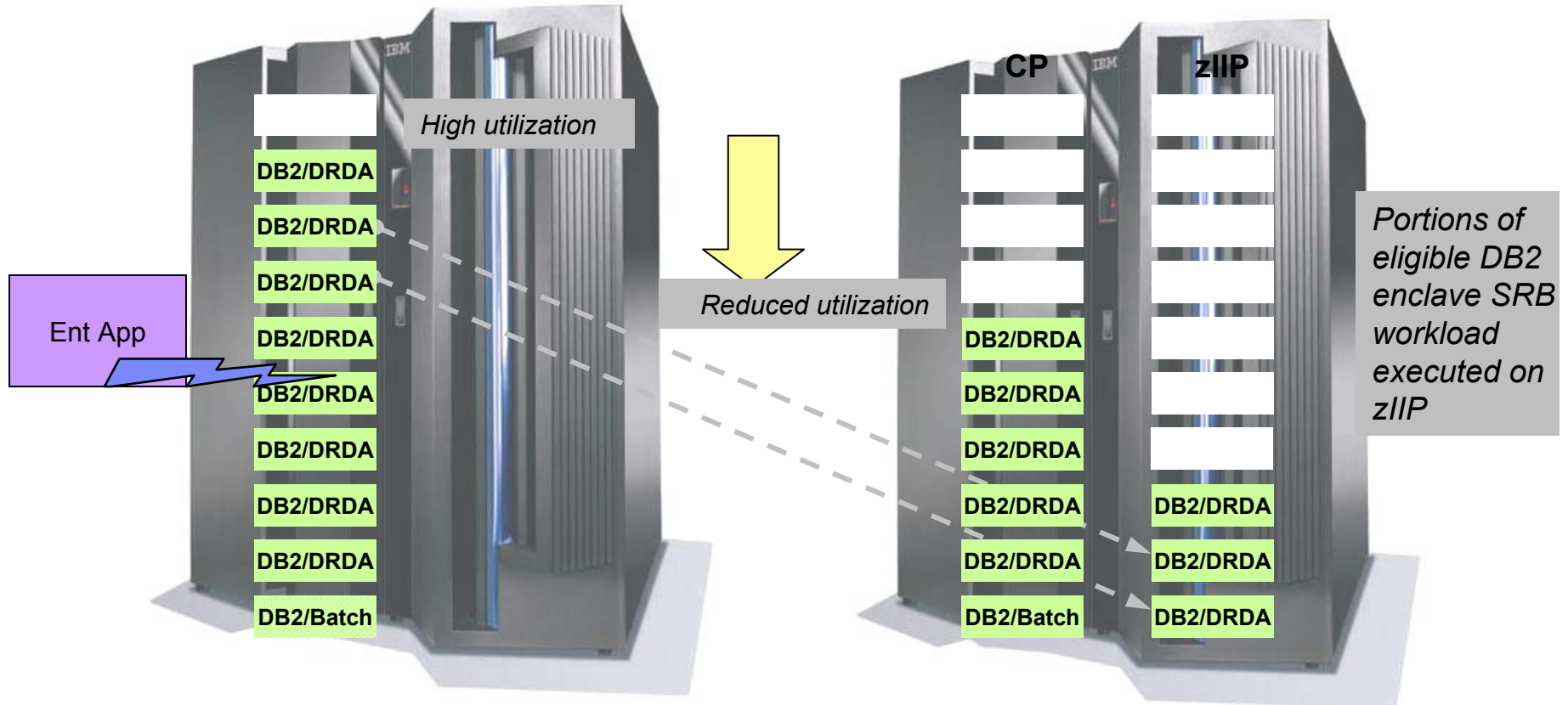
What are SRB Enclaves

- z/OS dispatches work in either TCB (Task Control Block) mode or SRB (Supervisor Request Block) mode. DB2 parallel tasks use SRB mode and are assigned the same importance as the originating address space.
- Preemptable enclaves are used to do the work on behalf of the originating TCB or SRB address space. Enclaves are grouped by common characteristics and service requests and since they are preemptable, the z/OS dispatcher (and WLM) can interrupt these tasks for more important ones (i.e. manage a transaction end-to end). There are two types of preemptable SRBs: client SRBs and enclave SRBs.
- If the DB2 for z/OS V8 request is coming in over distributed (i.e. DRDA over TCP/IP) then most of the work is executed in enclave SRBs.
- If the request is coming over local / native connection, then the work is dispatched between TCBs, client SRBs, and enclave SRBs (star schema parallel queries and some utility index maintenance now use enclave SRBs)

So..... regarding the zIIP, only the enclave SRB work (not the client SRB work or non-preemptable SRB work or TCB work) is eligible to be redirected to the zIIP. DB2 V8 knows how its work is dispatched and directs z/OS 1.6 or later to dispatch (redirect) a portion of the eligible work to the zIIP.

DRDA Example

Enterprise Applications that access DB2 for z/OS V8 via DRDA over a TCP/IP connection will have portions of these SQL requests directed to the zIIP



For illustrative purposes only
Actual workload redirects may vary

DB2 V8 zIIP re-direct enabling APAR:
PK18454 for DRDA - Available

DB2 Utilities

- Only portions of DB2 utility (LOAD, REORG, & REBUILD) processing related to index maintenance
 - Amount of work eligible to re-direct onto a zIIP, depends on:
 - How many indexes are defined on the table
 - How many partitions are in the table
 - If data compression is being used
 - And possibly other factors
 - Less re-direct is expected with:
 - Tables with fewer indexes
 - Fewer partitions
 - Compression used
 - More re-direct is expected with:
 - Tables with many indexes or many partitions
- DB2 V8 zIIP re-direct enabling APAR :
 - PK19920 for Utilities
 - Available

Star Schema

- Only Star Schema query parallel processing
 - A portion of the child task processing is eligible for re-direct to a zIIP
 - Main tasks coming in via DRDA via TCP/IP can take advantage of the DRDA use of zIIP.
- The combined child & main tasks coming in through DRDA via TCP/IP can increase the amount of processing eligible for zIIP.
- Longer running queries see higher benefit.
- DB2 V8 zIIP re-direct enabling APAR :
 - PK19921 for Star Schema
 - Available
 - PK27578 for more function
 - zIIP exploitation for all CP parallel queries
 - Available

What are the Tuning Options

- There are no external tuning options
- Different from the zAAP implementation
 - IFACROSSOVER
 - IFAHONORPRIORITY
- Internal tuning options
 - zIIP Crossover = NO
 - zIIP Honorpriority = YES
 - Needs help dispatcher
 - General purpose CPs only help if zIIP asks for help

RMF Report Updates

- RMF Monitor I CPU Activity Report

C P U A C T I V I T Y

z/OS V1R7

SYSTEM ID SYSC

START 07/09/20

RPT VERSION V1R7 RMF

END 07/09/20

CPU 2094 MODEL 750 H/W MODEL S54

---CPU---	ONLINE TIME	LPAR BUSY	MVS BUSY	CPU SERIAL	I/O TOTAL
NUM TYPE PERCENTAGE	TIME PERC	TIME PERC	TIME PERC	NUMBER	INTERRUPT
1 GP	100.00	51.84	51.84	07B10E	2612
2 GP	100.00	51.30	51.30	07B10E	2199
3 GP	100.00	51.49	51.49	07B10E	2374
4 GP	100.00	46.90	46.90	07B10E	54.87
CP TOTAL/AVERAGE		50.38	50.38		7240
5 AAP	100.00	57.05	57.05	07B10E	
AAP AVERAGE		57.05	57.05		
4 IIP	100.00	58.45	58.45	07B10E	
7 IIP	100.00	60.93	60.93	07B10E	
IIP AVERAGE		59.69	59.69		

GP CPs

zAAP CPs

zIIP CPs

zAAPs
And zIIPs
Do NOT
Process
Interrupts

RMF Report Updates

- RMF Monitor I Partition Report

START 07/07/2006-16.44.00 INTERVAL 000.11.59
 END 07/07/2006-16.56.00 CYCLE 1.000 SECONDS

Each processor type
 now identified as to
 type of functionality

NUMBER OF PHYSICAL PROCESSORS	54
CP	50
AAP	2
IFL	0
ICF	0
IIP	2

----- PARTITION DATA -----							-- LOGICAL	
NAME	S	WGT	DEF	ACT	DEF	WLM%	NUM	PROCESSOR-
								TYPE
SOSP7	A	DED	0	184		0.0	4.0	CP
SOSP16	A	DED	0	92		0.0	2	CP
SOSP16	A	DED					1	AAP
SOSP7	A	DED					2	IIP

Each processor type
 In separate pool on
 System z9

RMF Report Updates

- Workload Activity Report

REPORT BY: POLICY=WLMPOL WORKLOAD=WAS_WKL SERVICE CLASS=CISDDF RESOURCE GROUP=*NONE
 CRITICAL =NONE

TRANSACTIONS	TRANS-TIME	HHH.MM.SS.TTT	--DASD	I/O--	---SERVICE---	SERVICE TIMES	---APPL %---
AVG 37.78	ACTUAL	19	SSCHRT	11030	IOC	CPU 2228.5	CP 156.64
MPL 37.78	EXECUTION	19	RESP	2.3	CPU	59827K	AAPCP 0.00
ENDED 1372389	QUEUED	0	CONN	0.3	MSO	0	IIPCP 14.29
END/S 1906.10	R/S AFFIN	0	DISC	1.8	SRB	0	AAP 0.00
#SWAPS 0	INELIGIBLE	0	Q+PEND	0.2	TOT	59827K	IIP 152.88
EXCTD 0	CONVERSION	0	IOSQ	0.0	/SEC	83093	
AVG ENC 37.78	STD DEV	135				IIP 1100.7	
REM ENC 0.00					ABSRPTN	2200	
MS ENC 0.00					TRX SERV	2200	

GOAL: RESPONSE TIME 000.00.00.020 AVG

Service class should ONLY run enclave work, these numbers should be the same

TCB seconds and zIIP seconds – zIIP seconds are included in TCB seconds
 Rest of these fields should be 0 if all work is enclave SRB work

% of a zIIP CP consumed

% of a GP CP used by work eligible to run on a zIIP

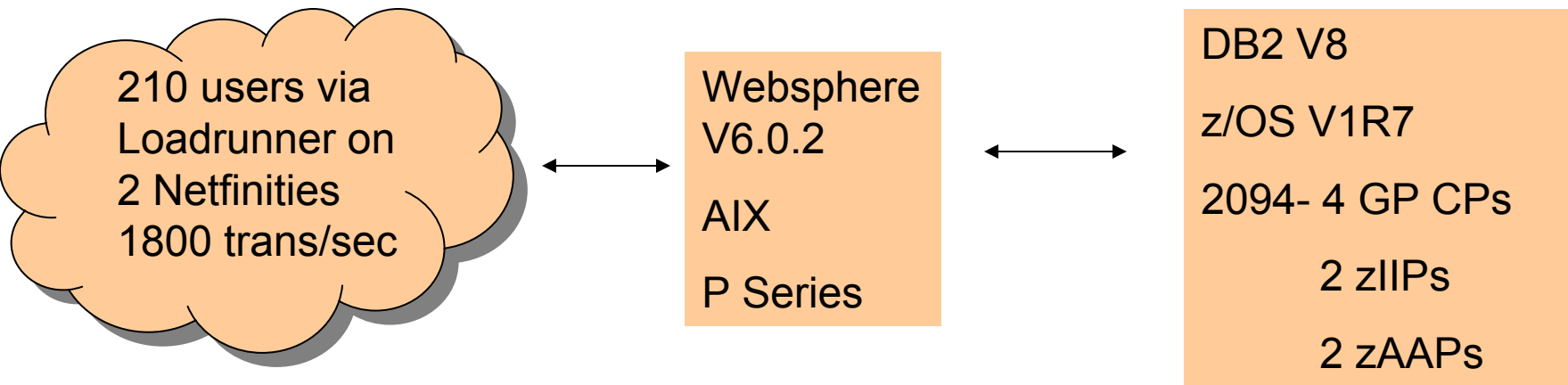
zIIP delay samples

zIIP using samples

RESPONSE TIME EX	PERF	AVG	-----	USING%	-----	EXECUTION DELAYS	%					
SYSTEM	HHH.MM.SS.TTT	VEL%	INDX	ADRSP	CPU	AAP	IIP	I/O	TOT	IIP	I/O	CPU
SYSC	000.00.00.019	29.3	1.0	37.8	2.8	0.0	2.8	7.8	14.6	12.9	1.0	0.6

Measurements

Environment - DRDA



DRDA Runs

- 4 GP CPs with 0 zIIPs to build the baseline
 - Identify zIIP eligible work using the RMF Post Processor Report
 - Validate workload can drive more than 1 zIIP
- 4 GP CPs with 1 zIIP
 - Validate work moved to the zIIP
 - IIP field on the RMF Workload Activity Report
 - Reduce utilization of the GP CPs
 - More work than 1 zIIP can handle
 - zIIP should 'ask for help'
 - IIPCP field on the RMF Workload Activity Report
- 4 GP CPs with 2 zIIPs
 - Validate more work is running on the zIIPs
 - Increase on IIP field on the RMF Workload Activity Report
 - Reduced IIPCP field on the RMF Workload Activity Report
 - Reduce utilization of the GP CPs
 - Less 'Needs Help' for zIIPs

DRDA Run # 2

1 zIIP – Exercise zIIP plus ‘Needs Help Dispatcher’

C P U A C T I V I T Y

z/OS V1R7				SYSTEM ID SYSC		START 07/07/200	
				RPT VERSION V1R7 RMF		END 07/07/200	
CPU	2094	MODEL 750	H/W MODEL	S54			
---	CPU---	ONLINE TIME	LPAR BUSY	MVS BUSY	CPU SERIAL	I/O TOTAL	
NUM	TYPE	PERCENTAGE	TIME PERC	TIME PERC	NUMBER	INTERRUPT	
0	CP	100.00	82.10	82.10	07B10E	8319	
1	CP	100.00	68.34	68.34	07B10E	195.6	
2	CP	100.00	80.01	80.01	07B10E	6559	
3	CP	100.00	68.20	68.20	07B10E	239.6	
CP	TOTAL/AVERAGE		74.66	74.66		15313	
4	IIP	100.00	93.46	93.46	07B10E		
IIP	AVERAGE		93.46	93.46			

REPORT BY: POLICY=WLMPOL WORKLOAD=WAS_WKL SERVICE CLASS=CISDDF RESOURCE GROUP=*NONE
 CRITICAL =NONE

TRANSACTIONS	TRANS-TIME	HHH.MM.SS.TTT	--DASD	I/O--	---	SERVICE----	SERVICE TIMES	---	APPL	%---
AVG	39.21	ACTUAL	21	SSCHRT	10965	IOC	0	CPU	1905.9	CP 232.70
MPL	39.21	EXECUTION	20	RESP	2.3	CPU	51166K	SRB	0.0	AAPCP 0.00
ENDED	1120884	QUEUED	0	CONN	0.3	MSO	0	RCT	0.0	IIPCP 83.26
END/S	1868.15	R/S AFFIN	0	DISC	1.8	SRB	0	IIT	0.0	
#SWAPS	0	INELIGIBLE	0	Q+PEND	0.2	TOT	51166K	HST	0.0	AAP 0.00
EXCTD	0	CONVERSION	0	IOSQ	0.0	/SEC	85276	AAP	0.0	IIP 84.96
AVG ENC	39.21	STD DEV	170					IIP	509.7	

DRDA Run # 3

2 zIIPs – Exercise zIIP reduced ‘Needs Help Dispatcher’

C P U A C T I V I T Y

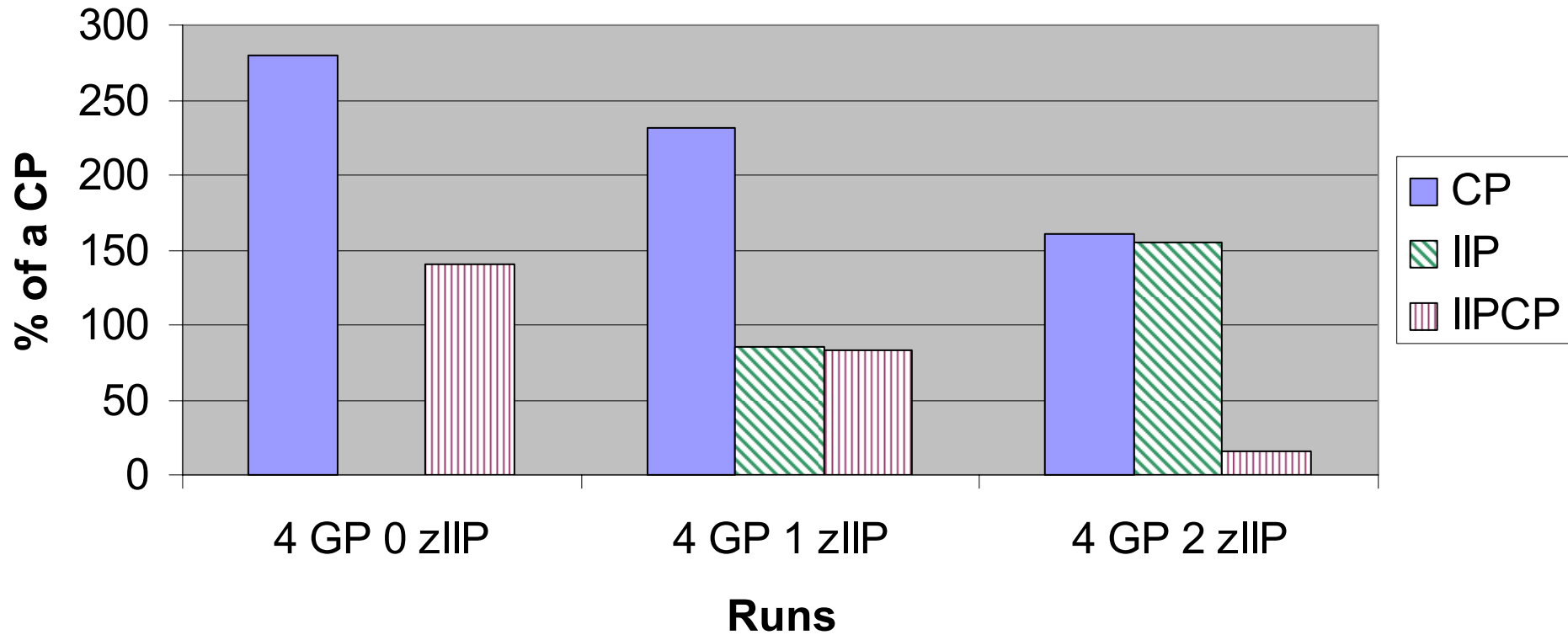
z/OS V1R7				SYSTEM ID SYSC		START 07/07/20	
				RPT VERSION V1R7 RMF		END 07/07/20	
CPU	2094	MODEL 750	H/W MODEL	S54			
---	CPU---	ONLINE TIME	LPAR BUSY	MVS BUSY	CPU SERIAL	I/O TOTAL	
NUM	TYPE	PERCENTAGE	TIME PERC	TIME PERC	NUMBER	INTERRUPT R	
0	CP	100.00	63.76	63.76	07B10E	5923	
1	CP	100.00	57.48	57.48	07B10E	4697	
2	CP	100.00	60.67	60.67	07B10E	4971	
3	CP	100.00	40.95	40.95	07B10E	119.3	
CP	TOTAL/AVERAGE		55.72	55.72		15710	
4	IIP	100.00	81.11	81.11	07B10E		
7	IIP	100.00	88.52	88.52	07B10E		
IIP	AVERAGE		84.81	84.81			

REPORT BY: POLICY=WLMPOL WORKLOAD=WAS_WKL SERVICE CLASS=CISDDF RESOURCE GROUP=*NONE
 CRITICAL =NONE

TRANSACTIONS	TRANS-TIME	HHH.MM.SS.TTT	--DASD	I/O--	---SERVICE----	SERVICE TIMES	---APPL %---
AVG	37.87	ACTUAL	19	SSCHRT	11070	IOC 0	CPU 1895.4 CP 160.73
MPL	37.87	EXECUTION	19	RESP	2.2	CPU 50884K	SRB 0.0 AAPCP 0.00
ENDED	1149619	QUEUED	0	CONN	0.3	MSO 0	RCT 0.0 IIPCP 15.57
END/S	1916.04	R/S AFFIN	0	DISC	1.8	SRB 0	IIT 0.0
#SWAPS	0	INELIGIBLE	0	Q+PEND	0.2	TOT 50884K	HST 0.0 AAP 0.00
EXCTD	0	CONVERSION	0	IOSQ	0.0	/SEC 84807	AAP 0.0 IIP 155.18
AVG ENC	37.87	STD DEV	122			IIP 931.1	

DRDA Summary

DRDA zIIP Measurements



Load Utility Runs

- 4 GP CPs with 0 zIIPs to build the baseline
 - Identify zIIP eligible work using the RMF Post Processor Report
- 4 GP CPs with 1 zIIP
 - Validate work moved to the zIIP
 - IIP field on the RMF Workload Activity Report
 - Reduce utilization of the GP CPs
 - Validate the projection of work moving to the zIIP
- 4 GP CPs with 2 zIIPs
 - Determine what impact the second zIIP might have

Load Utility Run # 1

Base Case

- CPU activity report not used due to job start and end not on interval boundaries
- Instead CPU times used for comparisons

TRANSACTIONS	TRANS-TIME	HHH.MM.SS.TTT	--DASD	I/O--	---SERVICE----	SERVICE TIMES	---APPL %---
AVG	0.94	ACTUAL	1.17.456	SSCHRT	16.4	IOC 2298	CP 99.57
MPL	0.94	EXECUTION	1.17.274	RESP	1.0	CPU 6412K	AAPCP 0.00
ENDED	3	QUEUED	181	CONN	0.8	MSO 4123M	IIPCP 13.33
END/S	0.01	R/S AFFIN	0	DISC	0.0	SRB 2818	IIT 0.00
#SWAPS	0	INELIGIBLE	0	Q+PEND	0.1	TOT 4129M	AAP 0.00
EXCTD	0	CONVERSION	0	IOSQ	0.0	/SEC 17205K	IIP 0.00
AVG ENC	0.00	STD DEV	2.06.561			IIP 0.00	

Utilities use dependent enclave SRBs versus independent enclaves.

Work continues in same period and service class. No additional WLM classification required

Job elapsed time

Total CPU time in seconds

$$\text{IIPCP Time} = 238.8 / 99.57 * 13.33 = 32 \text{ seconds}$$

- Appears less than 1 zIIP should be needed for this workload

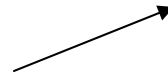
Load Utility Run # 2

1 zIIP

REPORT BY: POLICY=WLMPOL WORKLOAD=BAT_WKL SERVICE CLASS=DB2LOAD RESOURCE GROUP=*NONE
 CRITICAL =NONE

TRANSACTIONS	TRANS-TIME	HHH.MM.SS.TTT	--DASD	I/O--	---	SERVICE----	SERVICE	TIMES	---	APPL	%---
AVG	0.95	ACTUAL	1.16.920	SSCHRT	17.4	IOC	2429	CPU	243.5	CP	88.14
MPL	0.95	EXECUTION	1.16.532	RESP	0.9	CPU	6538K	SRB	0.1	AAPCP	0.00
ENDED	3	QUEUED	387	CONN	0.8	MSO	4122M	RCT	0.0	IIPCP	0.00
END/S	0.01	R/S AFFIN	0	DISC	0.0	SRB	2968	IIT	0.0		
#SWAPS	1	INELIGIBLE	0	Q+PEND	0.1	TOT	4128M	HST	0.0	AAP	0.00
EXCTD	0	CONVERSION	0	IOSQ	0.0	/SEC	17201K	AAP	0.0	IIP	13.38
AVG ENC	0.00	STD DEV	2.06.896					IIP	32.1		

Actual zIIP CPU seconds



Projected zIIP CPU seconds from run # 1 → IIPCP Time = $238.8 / .9957 * 13.33 = 32 \text{ seconds}$

- 13% of the load time was redirected to the zIIP
 - Function of database characteristics
 - 8 indexes, 20 partitions
 - Your mileage may vary

Load Utility Run # 3

2 zIIP

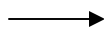
REPORT BY: POLICY=WLMPOL WORKLOAD=BAT_WKL SERVICE CLASS=DB2LOAD RESOURCE GROUP=*NONE
 CRITICAL =NONE

TRANSACTIONS	TRANS-TIME	HHH.MM.SS.TTT	--DASD	I/O--	---	SERVICE----	SERVICE	TIMES	---	APPL	%---
AVG	0.69	ACTUAL	1.23.010	SSCHRT	11.6	IOC	2429	CPU	242.9	CP	58.48
MPL	0.69	EXECUTION	1.22.546	RESP	0.9	CPU	6520K	SRB	0.1	AAPCP	0.00
ENDED	3	QUEUED	464	CONN	0.8	MSO	4120M	RCT	0.0	IIPCP	0.00
END/S	0.01	R/S AFFIN	0	DISC	0.0	SRB	2935	IIT	0.0		
#SWAPS	3	INELIGIBLE	0	Q+PEND	0.1	TOT	4127M	HST	0.0	AAP	0.00
EXCTD	0	CONVERSION	0	IOSQ	0.0	/SEC	11463K	AAP	0.0	IIP	9.02
AVG ENC	0.00	STD DEV	2.16.049					IIP	32.5		

Actual zIIP CPU seconds



Projected zIIP CPU seconds from run # 1

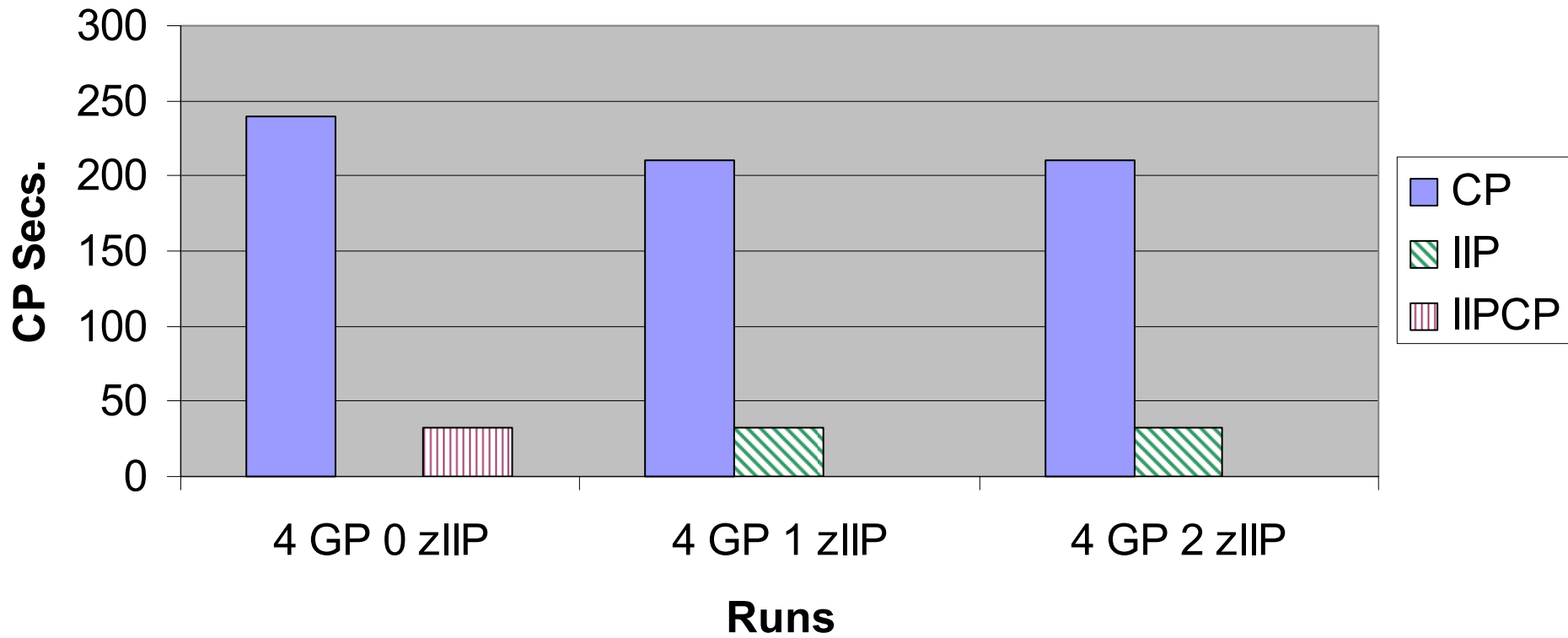


IIPCP Time = $238.8 / .9957 * 13.33 = 32 \text{ seconds}$

- 13% of the load time was redirected to the zIIP
 - Function of database characteristics
 - 8 indexes, 20 partitions
 - Your mileage may vary

Load Summary

Load zIIP Measurements



Reorg Utility Runs

- 4 GP CPs with 0 zIIPs to build the baseline
 - Identify zIIP eligible work using the RMF Post Processor Report
- 4 GP CPs with 1 zIIP
 - Validate work moved to the zIIP
 - IIP field on the RMF Workload Activity Report
 - Reduce utilization of the GP CPs
 - Validate the projection of work moving to the zIIP
- 4 GP CPs with 2 zIIPs
 - Determine what impact the second zIIP might have

Reorg Utility Run # 1

Base Case

- CPU activity report not used due to job start and end not on interval boundaries
- Instead CPU times used for comparisons

TRANSACTIONS	TRANS-TIME	HHH.MM.SS.TTT	--DASD	I/O--	---SERVICE----	SERVICE TIMES	---APPL %---
AVG	0.88	ACTUAL	3.36.918	SSCHRT	106.3	IOC 12630	CPU 164.9 CP 68.88
MPL	0.88	EXECUTION	3.35.824	RESP	2.5	CPU 4427K	AAPCP 0.00
ENDED	1	QUEUED	1.093	CONN	2.3	MSO 9454K	IIPCP 39.22
END/S	0.00	R/S AFFIN	0	DISC	0.0	SRB 9288	IIT 0.1
#SWAPS	0	INELIGIBLE	0	Q+PEND	0.2	TOT 13903K	HST 0.0
EXCTD	0	CONVERSION	0	IOSQ	0.0	/SEC 57927	AAP 0.0
AVG ENC	0.00	STD DEV	0				IIP 0.0

Utilities use dependent enclave SRBs versus independent enclaves.

Work continues in same period and service class. Not additional WLM classification required

Job elapsed time

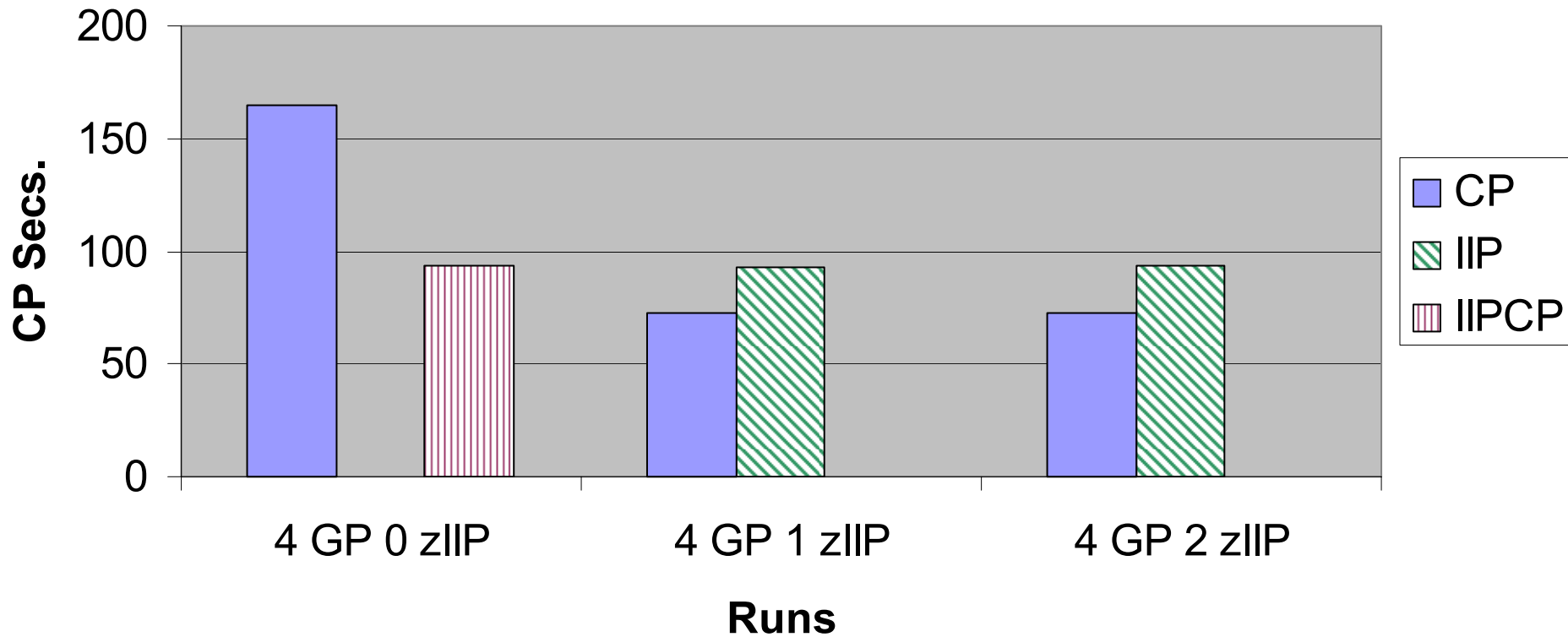
Total CPU time in seconds

$$\text{IIPCP Time} = 164.9 / 6888 * .3922 = 94 \text{ seconds}$$

- Appears less than 1 zIIP should be needed for this workload

Reorg Summary

Reorg zIIP Measurements



Summary

- The zIIP is a cost effective engine to support DB2 workloads on zSeries
 - Should allow zSeries to compete for workloads not considered in the past
- ProjectCPU provides accurate estimates for potential to redirect work to the zIIP
 - Requires z/OS v1.6 or higher and DB2 V8 plus zIIP PTFs
- No external tuning options reduces complexity of implementation
- Biggest inhibitor to exploitation is running a level of DB2 prior to V8

Questions

What About.....

