#SHAREorg

Experiences (zIIP-able to zIIP-ed)

Adrian Burke
IBM

March 13, 2012
Session 11078

Important Disclaimer

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY.

WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED.

IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE.

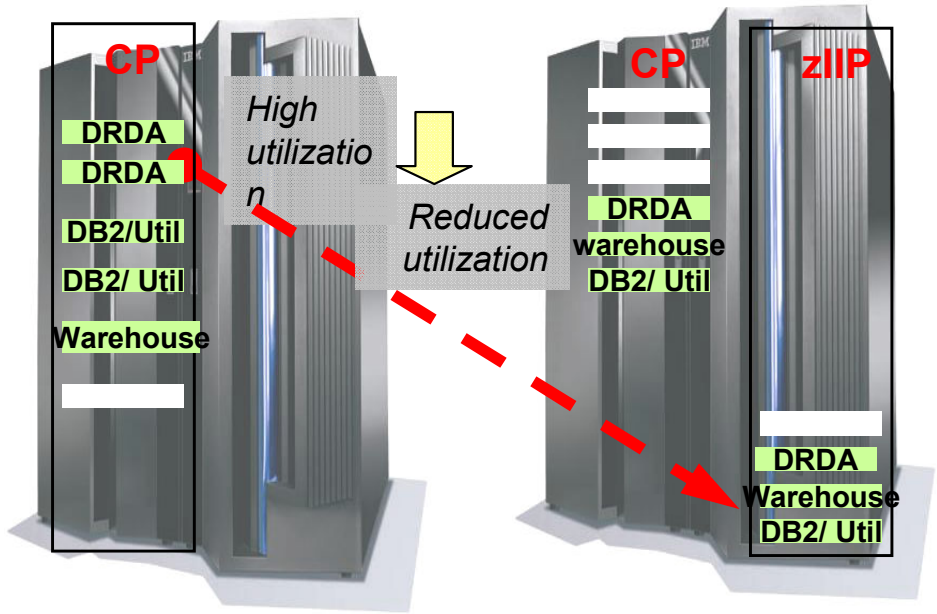
IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION.

NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, OR SHALL HAVE THE EFFECT OF:

- CREATING ANY WARRANTY OR REPRESENTATION FROM IBM (OR ITS AFFILIATES OR ITS OR THEIR SUPPLIERS AND/OR LICENSORS); OR
- ALTERING THE TERMS AND CONDITIONS OF THE APPLICABLE LICENSE AGREEMENT GOVERNING THE USE OF IBM SOFTWARE.

Agenda

- What is it?
- What's eligible?
- Where do I look?
  - Customer example
- What can I control?
- Recent enhancements

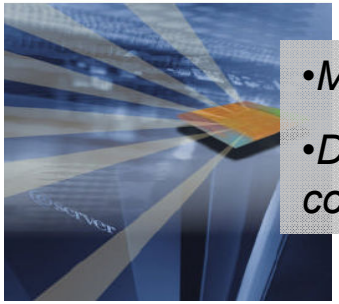


Adrian Burke
DB2 SWAT Team
agburke@us.ibm.com

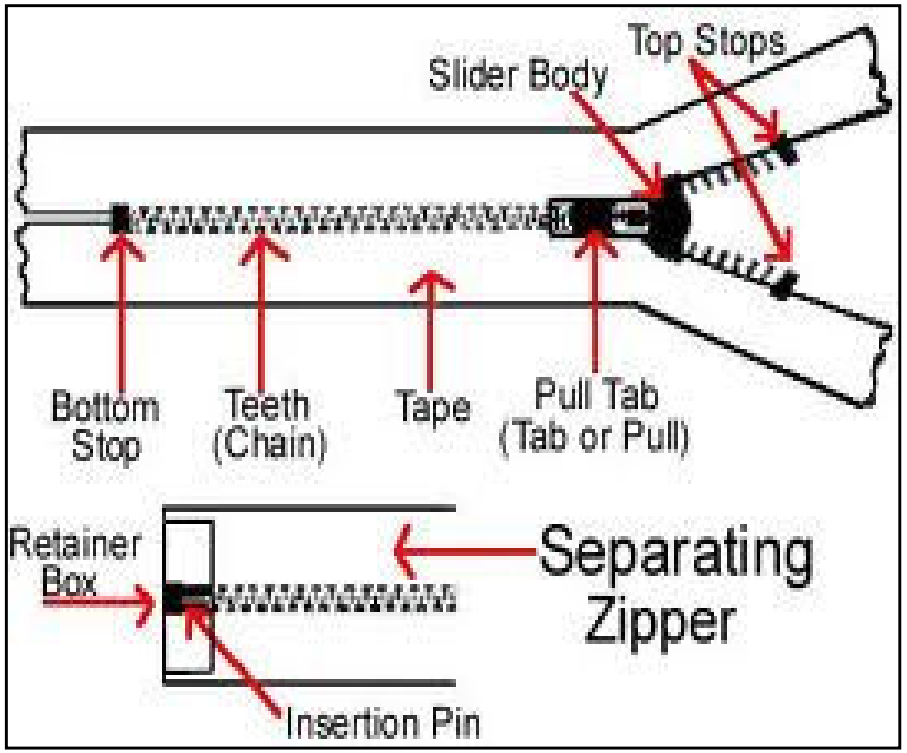
Certified for IBM Information Management software

SHARE in Atlanta 2012

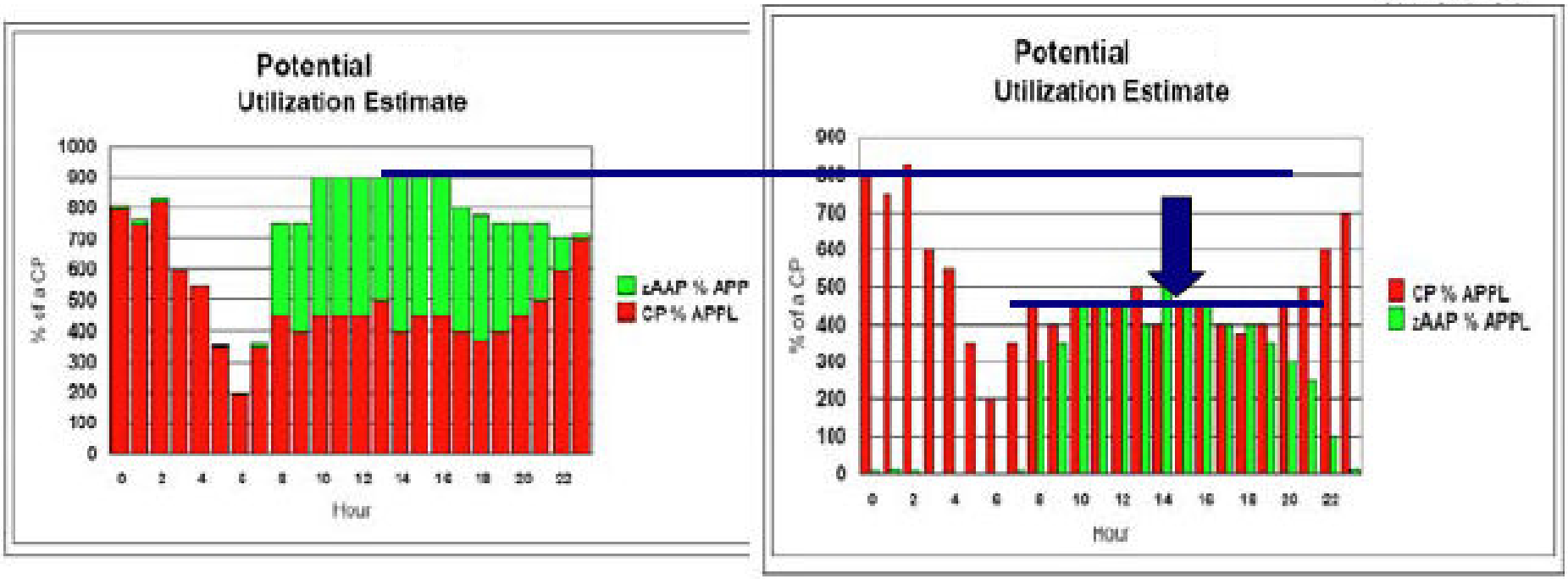
- Generally available in 2006 on the System z9 hardware
- IBM System z Integrated Information Processor (IBM zIIP)
- Quick jargon lesson
  - zIIP-able
    - Work that is zIIP eligible
  - zIIP-ed
    - Work that executed on a zIIP
  - Un-zIIP-ed
    - Eligible work that executed on a general CP



- *Microcode enabled*
- *Dispatcher controlled*



How can specialty engines help me?



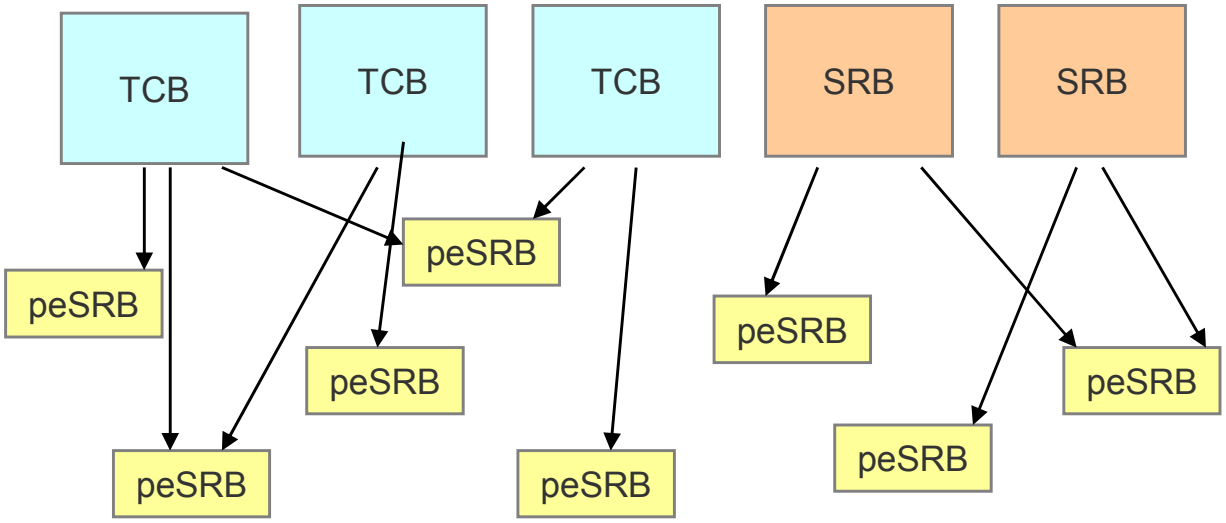
- Hardware costs: move work from GP to zIIP (zAAP), higher cost to lower cost processors, possibly postpone an upgrade
  - Specialty engines run at full rated speed of processor, so it could be the fastest one on the CEC
- Software costs: license and maintenance costs based on number of CPs in box
- BUT/AND…. it can also result in latent demand processing so processor utilization remains constant

zIIP Redirect Measurement Summary

- Companies noticing offload:
  - #3 credit reporter
    - is very happy it is estimated that they may save about $20M in ISV software savings in 3 years (50% workload from DDF on zIIP)
  - SAP customer running application servers on Linux on system z and DB2 for z/OS
    - 60% of MIPs related to DB2 workload offloaded
    - Roughly half of their MLC bill went away due to the zIIP

Work is dispatched

DB2 for z/OS



There are four types **of dispatchable units in z/OS:**

- **Preemptible Task Control Block (TCB)**
- **Non-preemptible Service Request Block (SRB)**
- **Preemptible Client Service Request Block (client SRB)**
- **Preemptible Enclave Service Request Block (enclave SRB)**

DBM1 SRBs (* means data sharing)

- **DB Writes/ Asynchronous I/O**
- **Memory management**
- **Prefetches**
- **Real time stats**
- ***castout***
- ***P-lock negotiation***
- ***SYSLGRNX***
- ***GBP checkpoints***
- **Backout preemptible (V10)**

DBM1 TCBs

- **Open/close**
- **Pre-format/ extend**
- **Statistics**
- **Full system contraction**

What are enclave SRBs?

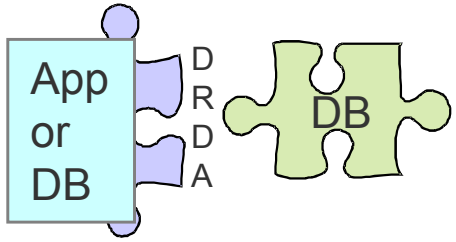
- z/OS dispatches DB2 work in either TCB, Client SRB, or Enclave SRB mode if request is local or an Enclave SRB (Service Request Block) mode if request is distributed. Under these modes of operation the parallel tasks are assigned the same importance as the originating address space.
- Preemptible enclaves are used to do the work on behalf of the originating TCB or SRB address space. Enclaves are grouped by common characteristics and service classes and since they are preemptible, the z/OS dispatcher (and WLM) can interrupt these tasks for more important ones (ie manage a transaction end-to-end). There are two types of preemptible SRBs: client SRBs and enclave SRBs.

- If the DB2 for z/OS request is coming in over distributed (ie DRDA over TCP/IP) then that work is executed in enclave SRBs.
  - only the enclave SRB work is eligible to be redirected to the zIIP.

- DB2 knows how its work is dispatched and directs z/OS to dispatch (redirect) a portion of the eligible work to the zIIP.

What is DRDA?

- DRDA = Distributed Relational Database Architecture
  - Developed by IBM
  - Enables relational data to be distributed among multiple platforms – 'any app to any db and any db to any db'. Applications and APIs accomplish the actual implementation



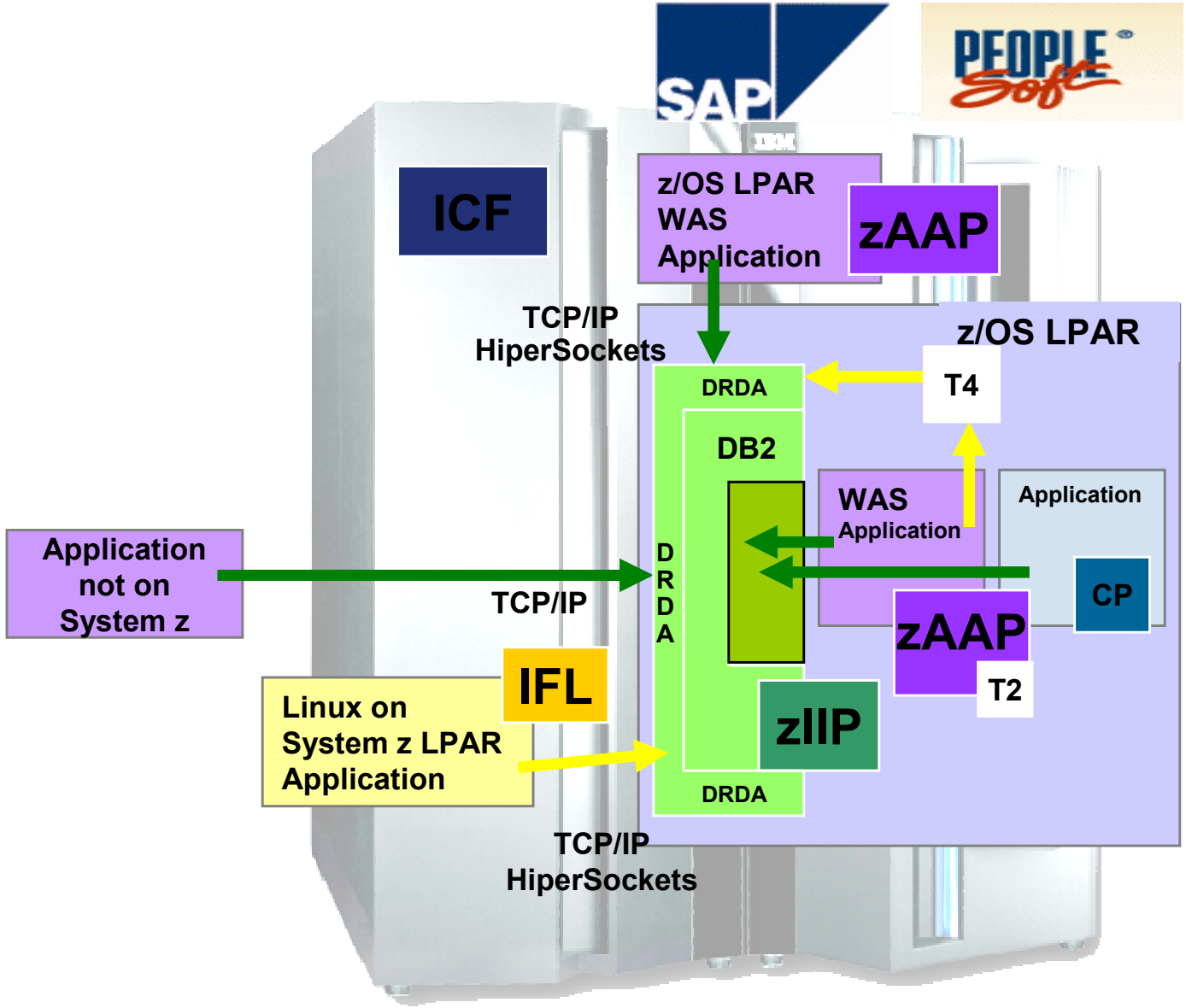
- DRDA is native to DB2 for z/OS. It reduces the need for additional gateway products that may affect performance and availability
  - Private Protocol, also IBM, was stabilized in V4 of DB2
- The Open Group adopted DRDA in 1998 as the open standard for database access interoperability (DB2, Informix, Oracle)
- DRDA can use TCP/IP or SNA as a network protocol to flow commands

So.... regarding the zIIP: if DB2 for z/OS workload comes over TCP/IP and is DRDA compliant, a portion of that DB2 workload is eligible to be redirected to the zIIP.

As of V9 SNA protocol incurs an **overhead**

And Private Protocol support is removed in V10

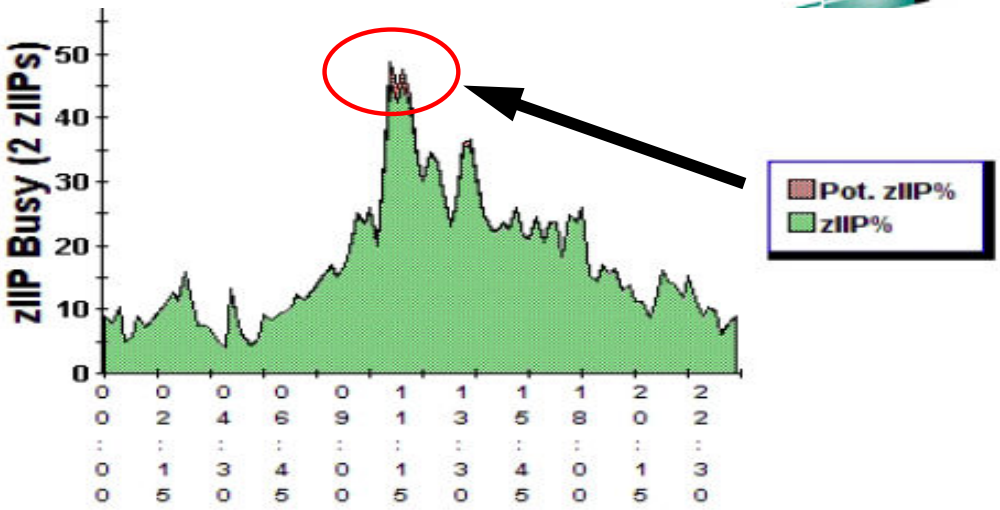
Specialty engines

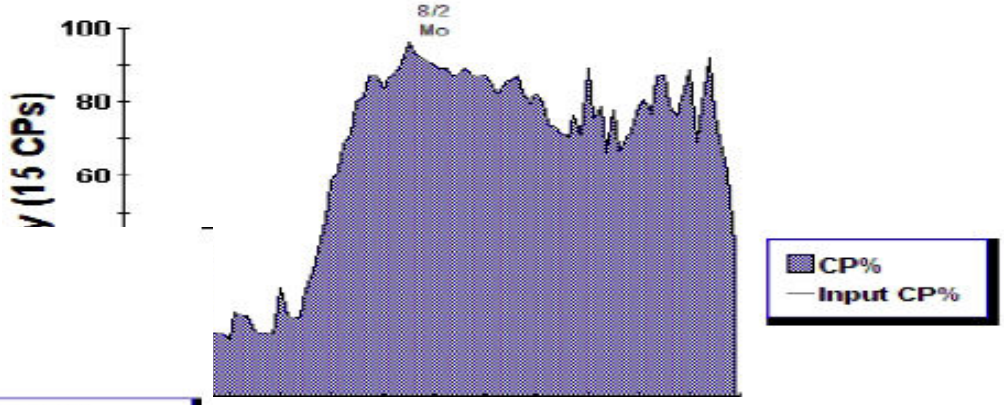


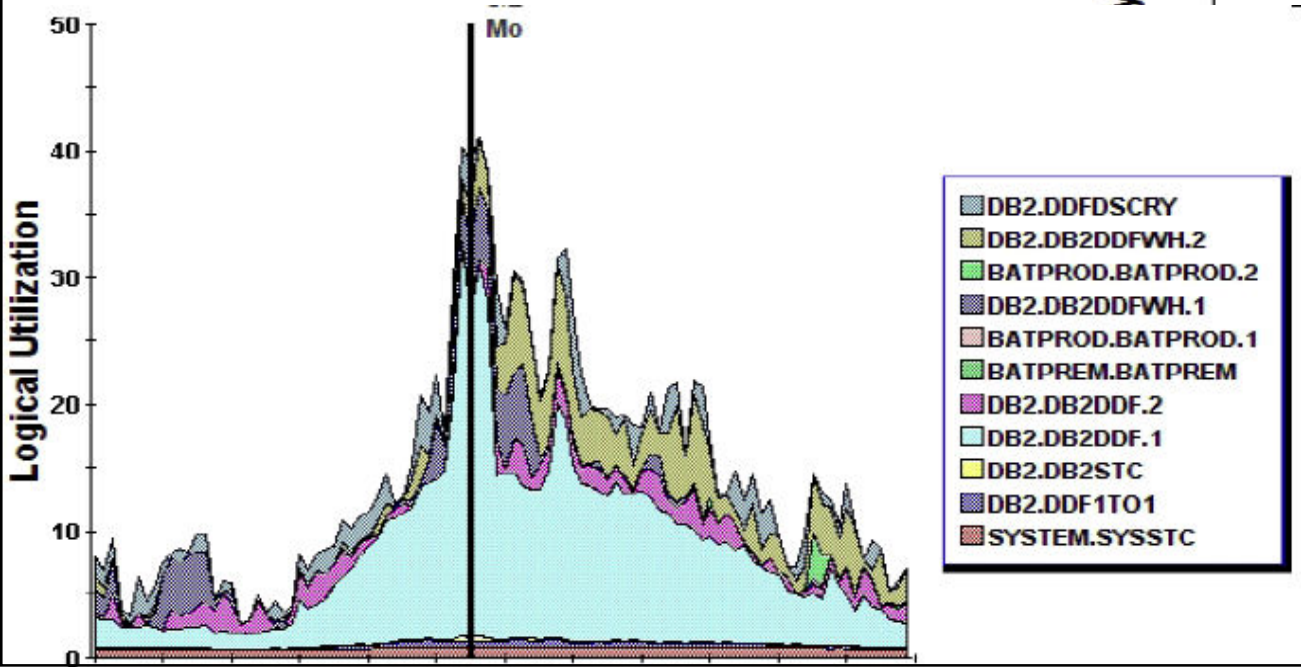
The IBM System z specialty engines can operate on the same machine together

How many should I have?

- PROJECTCPU=YES (IEAOPTx)
- zCP3000 study
  - Provided by IBM techline
    - Send in SMF 30's, 70's
  - Breakdown of eligible work
  - Overlay 4 hour peak
  - See collisions of workloads

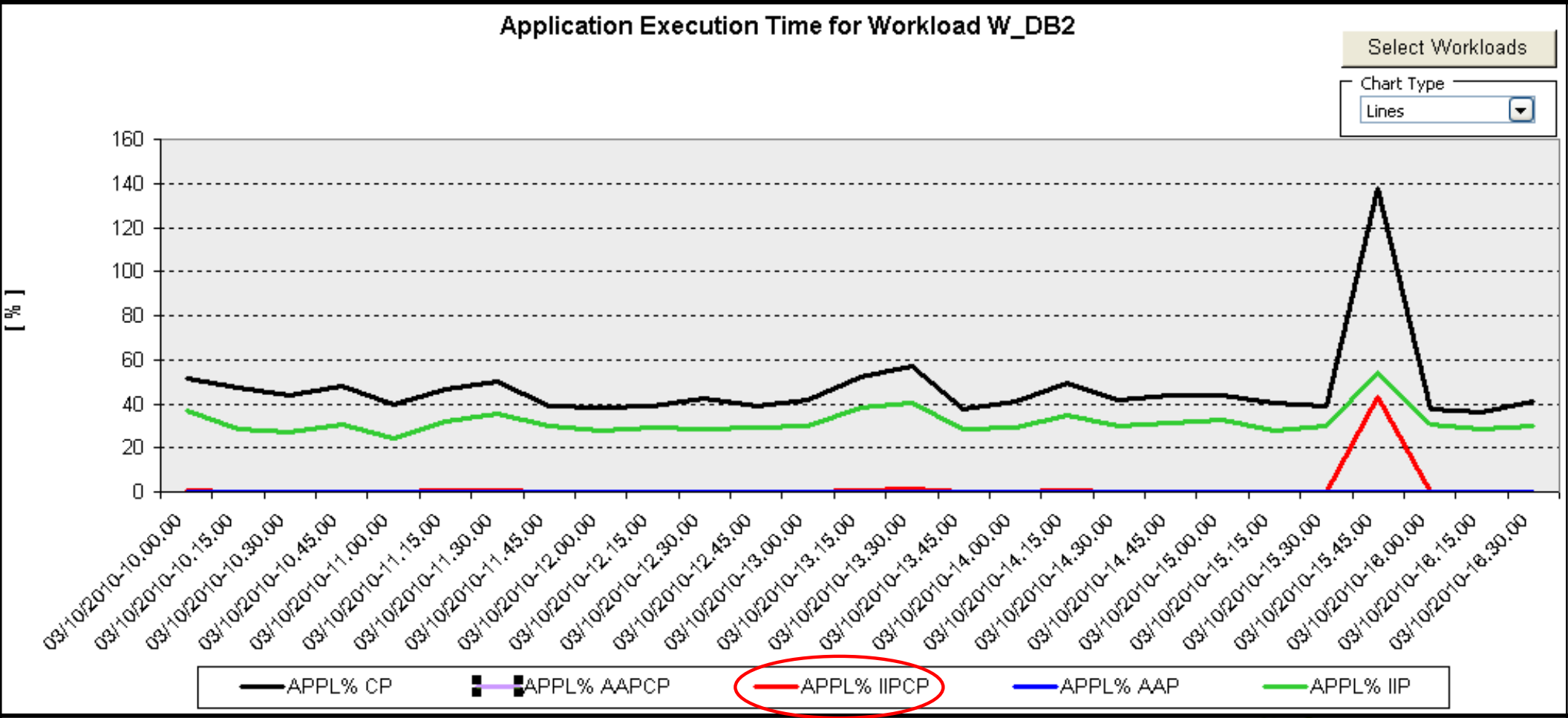






WLM and the zIIP

- How many zIIPs do you need (this scenario 12:1)
- Law of probability for many CPs vs. zIIPs



zIIP Eligible APAR II14219	Function *Read left to right all the way across 1)Function -> 1)Amount Redirected -> 1)Prerequisites	Amount Redirected	Prerequisites z/OS 1.8 – base feature z/OS1.9 – WLM weights on zIIPs
DB2 V8	1) Utilities 2) Distributed DRDA requests 3) Parallelism (star schema and parallel queries) 4) Result set of remote Stored procedures	1) Up to 60% in Lab measurements (see next slide for break down): more partitions and indexes= more redirect (BUILD and REBUILD phases of index maintenance go to zIIP) 2) Up to 60% in Lab measurements. 3) Portion of main task for remote calls, Portions of child tasks in both cases 4) Call, commit, result-set processing	1) UK15814 2) DRDA over TCP/IP – PM12256 3) zPARM CDSSRDEF=1, PARAMDEG =0 (>0 to limit the degree of parallelism) DEGREE ANY bind parameter and SET CURRENT DEGREE ANY at statement level. 4) N/A
DB2 9	1) All the offload in V8 plus the following 2) Distributed calls to Native Stored Procedures 3) XML parsing offloaded to zAAP and zIIP	1) Slightly less for Utilities due to CPU reduction for index processing in DB2 9 but added UNLOAD phase during REORG 2) Remote calls offload same percentage as remote DRDA requests 3) Up to 36% zAAP redirect in Lab measurements for XML LOAD utility. Up to 63% zIIP redirect in Lab measurements for XML INSERT via DRDA.	1) PM37622 2) No FENCED or EXTERNAL keywords, native SQL code 3) Z/OS 1.8
Other Processes	1) IPSec 2) Global Mirror for z/OS (formerly Extended Remote Copy) 3) HiperSockets for Large messages 4) DFSORT 5) zAAP on zIIP	1) Encryption processing, header processing and crypto validation (93% for bulk data movement) 2) Most System Data Mover processing 3) Handles large outbound messages (multiple channel paths given to SRBs) 4) Sorting of fixed length rows 5) zAAP eligible work can move to zIIP if no zAAP installed	1) z/OS 1.8 + UA34582 AND z/OS Communication Server PTF UK27062-63 2) z/OS 1.10, or 1.9 + UA39510, or 1.8 + UA39509 (zGM parmlib zIIPEnable) 3) z10 and z/OS 1.10 (GLOBALCONFIG ZIIP IQDIOMULTIWRITE) 4) PK85899 and PK85856 (z/OS 1.10) 5) z/OS 1.11 base or 1.9 or 1.10 w/ APAR OA27495

zIIP Eligible APAR II14219	Function *Read left to right all the way across 1) Function -> 1) Amount Redirected -> 1) Prerequisites	Amount Redirected	Prerequisites z/OS 1.8 – base feature z/OS1.9 – WLM weights on zIIPs
DB2 10	1) All of DB2 v8 and 9 offload++ 2) RUNSTATS 3) Prefetch and deferred write processing 4) Parallelism enhancements	1) BUILD phase, Native SQL procs, parallelism,, 60% DRDA requests 2) Basic RUNSTATS for table, NO Histogram, DSTATS, COLGROUP… BUT index stats almost all offloaded (not DPSIs) 3) 100% 4) Offload greater than 60%	1) DB2 10/ z/OS 1.10 2) Run RUNSTATS, no inline STATS 3) Shows up in DBM1 SRB time 4) V10 NFM with rebind
DB2 9	1) All the offload in V8 plus the following 2) Distributed calls to Native Stored Procedures 3) XML parsing offloaded to zAAP and zIIP	1) Slightly less for Utilities due to CPU reduction for index processing in DB2 9 but added UNLOAD phase during REORG 2) Remote calls offload same percentage as remote DRDA requests 3) Up to 36% zAAP redirect in Lab measurements for XML LOAD utility. Up to 63% zIIP redirect in Lab measurements for XML INSERT via DRDA.	1) PM37622 2) No FENCED or EXTERNAL keywords, native SQL code 3) Z/OS 1.8
Other Processes	1) IPSec 2) Global Mirror for z/OS (formerly Extended Remote Copy) 3) HiperSockets for Large messages 4) DFSORT…DB2SORT 5) zAAP on zIIP	1) Encryption processing, header processing and crypto validation (93% for bulk data movement) 2) Most System Data Mover processing 3) Handles large outbound messages (multiple channel paths given to SRBs) 4) Sorting of fixed length rows in DFSORT (10-40% Utility) / 10-20% for DB2SORT 5) zAAP eligible work can move to zIIP if no zAAP installed	1) z/OS 1.8 + UA34582 AND z/OS Communication Server PTF UK27062-63 2) z/OS 1.10, or 1.9 + UA39510, or 1.8 + UA39509 (zGM parmlib zIIPEnable) 3) z10 and z/OS 1.10 (GLOBALCONFIG ZIIP IQDIOMULTIWRITE) 4) PK85899 and PK85856 (z/OS 1.10) 5) z/OS 1.11 base or 1.9 or 1.10 w/ APAR OA27495

zIIP Eligibility cont.

```
INTERVAL = REORG                CPU (SEC)          = 0.258516
   LEVEL = UTILITY              ELAPSED TIME (SEC) = 5.078
```

```
INTERVAL = BUILD                CPU (SEC)          = 0.066509
   LEVEL = PHASE                ELAPSED TIME (SEC) = 3.900
```

- Utilities (Lab measurement with 4 CPs, 2 zIIPs)
  - 5 to 20% for Rebuild Index
  - 10 to 20% for Load or Reorg of a partition with one index only, or Load of entire table, or Reorg of entire tablespace
  - 40% for Rebuild Index of logical partition of non partitioning index
  - 40 to 50% for Reorg Index
  - 30 to 60% for Load or Reorg of a partition with more than one index
- OMEGAMON Accounting Long report in CPU section – CPU seconds normalized
  - SE CPU : Actual redirect
  - SECP CPU: zIIP eligible that executed on general CP or projection

  (IIP changed to SE (specialty engine) with APAR PK51045/ **N/A in DB2 10 [next slide]**)
- RMF Workload Activity Report APPL% value
  - SYS1.PARMLIB(IEAOPTxx) parameter - PROJECTCPU=YES (for projection with no zIIP)
    - CP=% executed on general CP
    - IIPCP=% zIIP eligible that executed on general CP or projection
    - IIP=% actual redirect to zIIP
- DFSORT with MSGICE256I DFSORT CODE IS ELIGIBLE TO USE ZIIP FOR THIS DB2 UTILITY RUN **(PK85899)**

Tivoli Omegamon DB2PE Accounting Report with Local Parallel Query zIIP Redirect

AVERAGE	APPL(CL.1)	DB2 (CL.2)
CP CPU TIME	19.373768	19.365788
AGENT	6.779348	6.771411
NONNESTED	6.779348	6.771411
STORED PRC	0.000000	0.000000
UDF	0.000000	0.000000
TRIGGER	0.000000	0.000000
PAR.TASKS	12.594420	12.594377
IIPCP CPU	2.813831	N/A
IIP CPU TIME	35.886951	35.886951

Chargeable CPU time.
Includes IIPCP CPU time.
Does not include IIP CPU time.
So 2.813831 included in 19.373768

IIP => SE after APAR PK51045

zIIP eligible but ran on CP, deprecated in V10
But being brought back through PM57206

CPU time on zIIP

Total zIIP eligible work % = 70% ((IIP +IIPCP) / (CP+IIP))
zIIP Redirect % = 65% ((IIP / (CP+IIP))
zIIP eligible but ran on CP = 5% ((IIPCP / (CP+IIP))

XML offload – ATS chart

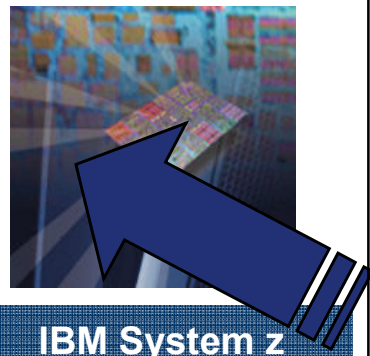
Workload	Examples	Available	Redirect	Requirements
z/OS XML System Services, non-validating parsing, – executing in TCB mode	- any SW using z/OS XML System Services parsing in TCB - EXAMPLE : local applications inserting/ saving XML data, and XML table loads on DB2 9	Sept, 2007 (with z/OS V1.9 GA)	100% of z/OS XML System Services parsing - eligible for zAAP	DB2 9 New Function Mode
z/OS 1.9 z/OS 1.8 (with OA20308) z/OS 1.7 (w/ OA16303 and OA20308)	- select XML Toolkit for z/OS V1.9 parsing workloads	May 9, 2008	same	XML Toolkit for z/OS V1.9 (5655-J51) with PTF for APAR OA22700
C API for z/OS XML System Services available with z/OS 1.9, and z/OS 1.7 & 1.8 with APAR OA18713	- Enterprise COBOL V4.1, using XMLPARSE option	Dec, 2007 (with COBOL V4.1 GA)	same	Enterprise COBOL V4.1 (5655-J51) z/OS V1.7- V1.9 with OA22777 (SW Announce 207-339).
z/OS XML System Services, non-validating parsing, executing in enclave SRB mode	- DB2 9 inserting/ saving XML data using DRDA via TCP/IP	Sept. 2006 (with GA of z/OS V1.8)	Same % as the zIIP-eligible work (DRDA)	DB2 9 New Function Mode
z/OS 1.8, z/OS 1.7 w/ OA16303	- any SW (including DB2 9) using z/OS XML System Services in enclave SRB mode	Soon (GA of z/OS V1.10)*	100% of z/OS XML System Services parsing eligible for zIIP	z/OS 1.9 and 1.8 (both with APAR OA22035) DB2 9 NFM
z/OS XML System Services with validating parsing, both enclave SRB and TCB modes.	-- any SW using z/OS XML System Services validating parsing	Soon (GA z/OS V1.10)*	100% of z/OS XML System Services validating parsing eligible for zAAP (TCB) or zIIP (enclave SRB)	z/OS V1.10 *
	- select XML Toolkit for z/OS V1.9 workloads	SOD*	TBD	TBD
Java-based XML parsing	- applications using Java-based XML parser in IBM SDK - any SW performing XML parsing/ processing in Java	Yes (with availability of zAAP)	100% of Java-based XML parsing eligible for zAAP	Any z/OS, System z processor with zAAP support.

XML on zIIP – ATS chart

DB2 9 XML invoked from	DRDA	Execution Mode	zAAP eligible?	zIIP eligible?
CICS/IMS/TSO	No	TCB	Yes	No
WAS z/OS JCC T2	No	TCB	Yes	No
SP / UDF / Trigger	No	TCB	Yes	No
Native SQL SP	No	TCB	Yes	No
CAF	No	TCB	Yes	No
RRS	No	TCB	Yes	No
Load (local) Data	No	TCB	Yes	No
Index Build for Load	No	SRB	No	Yes
WAS z/OS JCC T4	Yes	SRB	No	Yes
WAS Distrib JCC T4	Yes	SRB	No	Yes
Dist w/DB2 Connect	Yes	SRB	No	Yes
SP / UDF / Trigger	Yes	TCB	Yes	No
Native SQL SP	Yes	SRB	No	Yes

zAAP on zIIP

IBM System z9 Integrated Information Processor (IBM zIIP) 2006

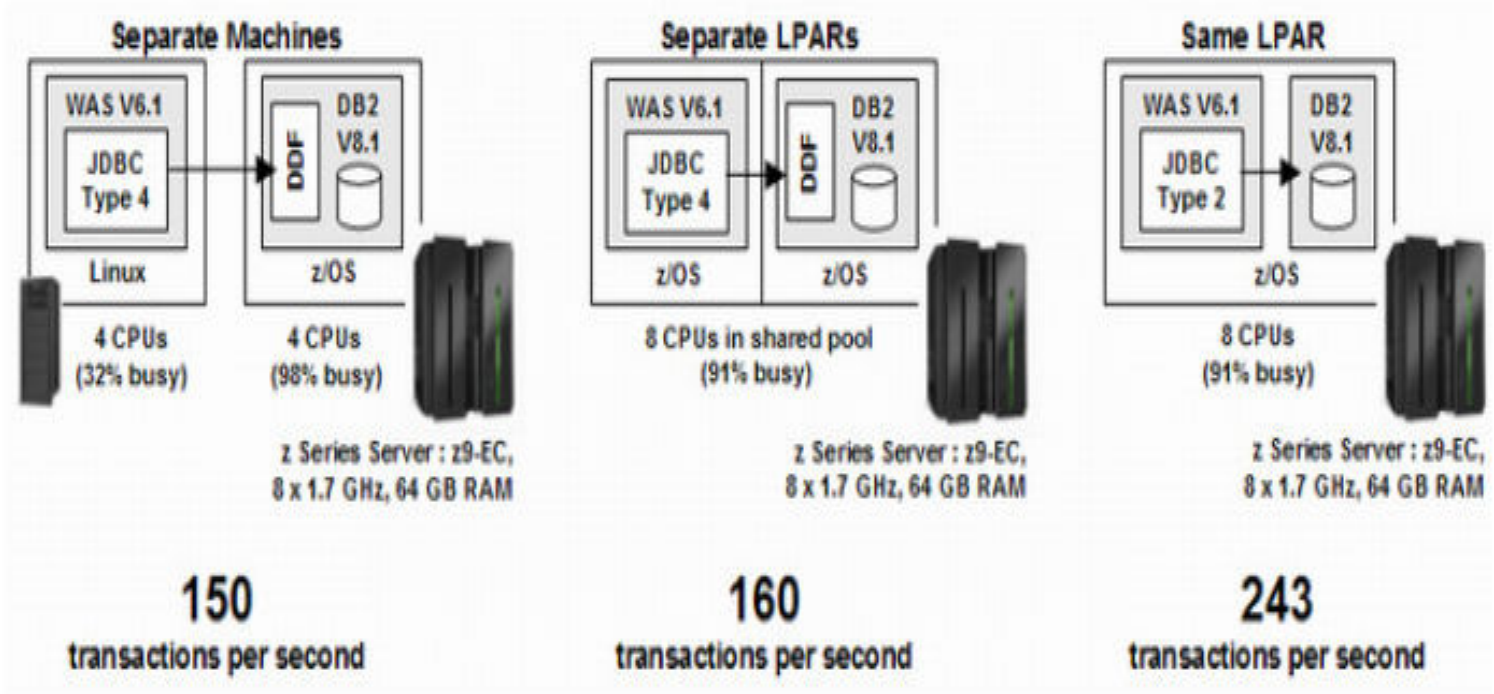


IBM System z Application Assist Processor (zAAP) 2004

- Allows zAAP eligible work to run on a zIIP (JVM, XML parsing)
- Native in z/OS 1.11
  - z/OS 1.9 and 1.10 APAR OA27495
  - ZAAPZIIP=yes in IEASYSxx parmlib
- IFAHONORPRIORITY now based on IIPHONORPRIORITY
- No IFAACROSSOVER
- Still a 1:1 zIIP:CP ratio
- zAAP on zIIP planning guide - http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD103548

T2 vs. T4 Connectivity and Specialty Engines

- Type 4 saved CP cost over 3 years ago
  - Even then T4 used 23% more total processing cycles (PTV8)
  - 62% less ITR (seen here)
  - 19% increase in elapsed time (PVT8)
  - Current V10 numbers show T2

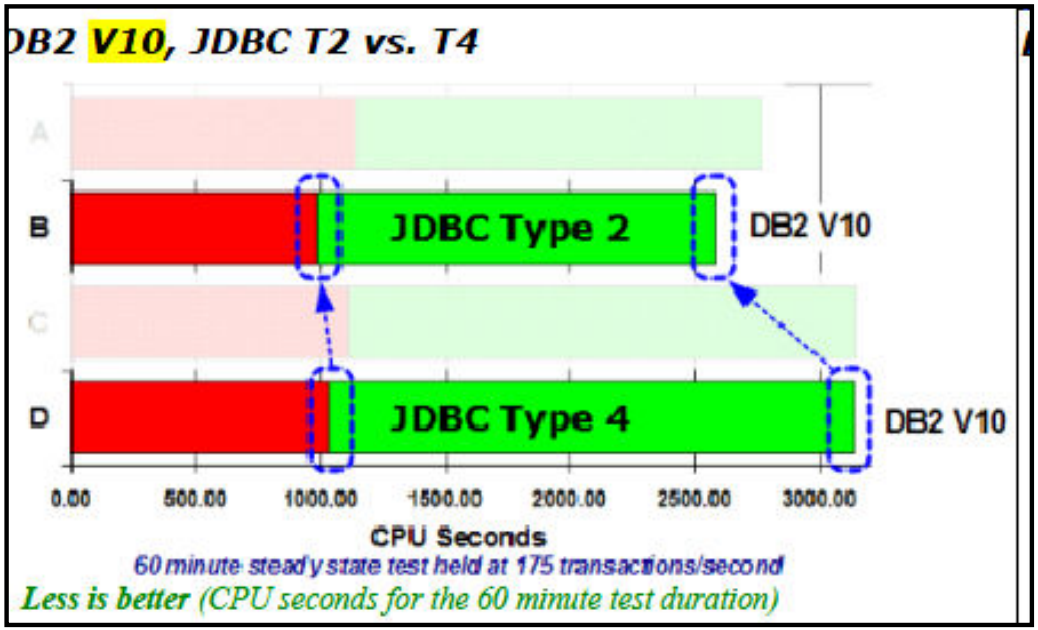


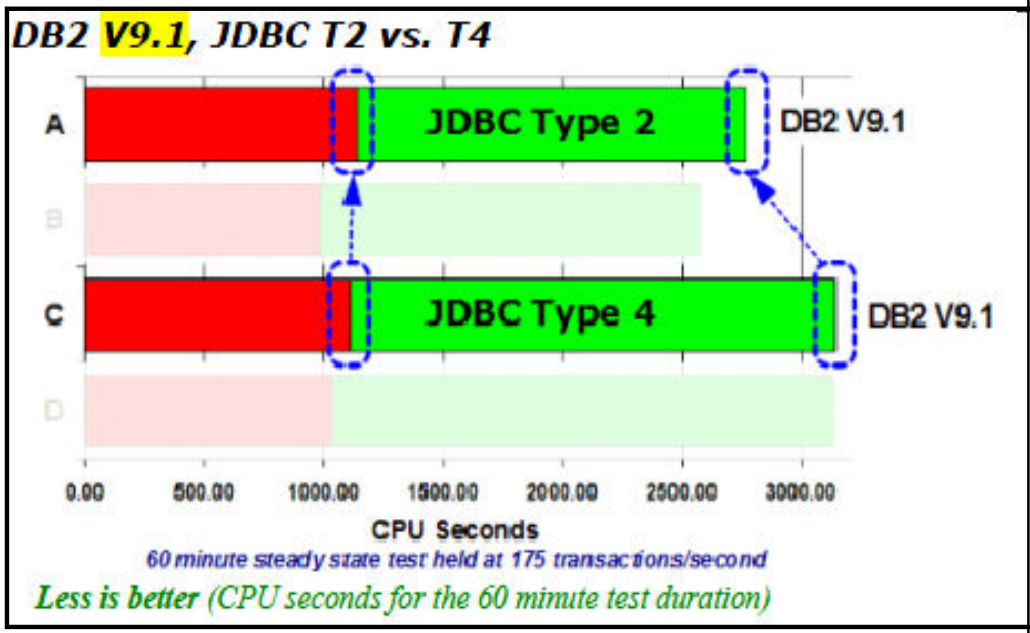
April 2009 Banking Scalability Study – WP101476 'The Value of Co-location'

http://www-03.ibm.com/support/techdocs/atsmastr.nsf/5cb5ed706d254a8186256c71006d2e0a/94f9d53f2e526489862575c5004b50b1/$FILE/WP101476%20-%20WebSphere%20zOS%20-%20The%20Value%20of%20Co-Location.pdf

http://www-03.ibm.com/support/techdocs/atsmastr.nsf/5cb5ed706d254a8186256c71006d2e0a/94f9d53f2e526489862575c5004b50b1/$FILE/WP101476-2%20-%20Value%20of%20Co-Location%20Update.pdf

Cost of deployment for type 2 vs. type 4 drivers





http://www-03.ibm.com/support/techdocs/atsmastr.nsf/5cb5ed706d254a8186256c71006d2e0a/94f9d53f2e526489862575c5004b50b1/$FILE/WP101476-2%20-%20Value%20of%20Co-Location%20Update.pdf

Type 2 and type 4 example from WAS on z/OS customer prior to improvements, NOT benchmarked

- The SQL profile was slightly different on these two days
  - Type 2 driver
    - 6.173 ms Class 2 CPU
    - 32.85 ms Class 2 ET
  - Type 4 driver
    - 4.775 ms Class 2 CPU
    - 63.5 ms Class 2 ET
  - Not Accounted for time
    - RRS workload had 9% Not Accounted for time.
    - DDF workload (type 4) had 27%

DRDA zIIP Redirect Summary

- Measured with Stored Procedure distributed workloads.
  - External stored procedure achieved 10% redirect
    - Stored Procedure Call, Results set and Commit processing eligible for zIIP redirect.
- DB2 9 Native SQL Procedure SQL processing is eligible for DRDA level of zIIP redirect (60% with 2010 maintenance)
- Parallel Query workload achieved expected redirect %
- No noticeable CPU overhead or elapsed time increase for zIIP redirect processing.

Stored Procedures with zIIPs

Language	Base Billable Cost	Billable Cost after zIIP and/or zAAP acceleration
COBOL stored proc	X (Baseline)	.88x
C stored proc	.95x	.83x
Remote SQLJ	1.78x	1.06x
SQLJ stored proc	1.21x	1.15x (zIIP + zAAP)
JDBC stored proc	2.11x	1.76x (zIIP + zAAP)
External SQL stored proc	1.62x	1.49x
Native SQL stored proc	1.14x	.65x

Asynchronous I/O (V10)

- In DB2 10 prefetch and deferred write are zIIP eligible
  - After PM30468 reported in DBM1 SRB time
  - Increase due to index I/O parallelism/ index list prefetch for disorganized indexes/ access path changes/ more dynamic prefetch in V9,V10

DB2 VERSION: V8 SCOPE: MEMBER TO: 09/10

---- HIGHLIGHTS ----

INTERVAL START	: 09/09/11 05:30:01.83	SAMPLING START:	09/09/11 05:30:01.83	TOTAL THREADS	: 90.00
INTERVAL END	: 09/10/11 05:00:02.70	SAMPLING END	: 09/10/11 05:00:02.70	TOTAL COMMITS	: 6328.8K
INTERVAL ELAPSED:	23:30:00.864709	OUTAGE ELAPSED:	0.000000	DATA SHARING MEMBER:	N/A

CPU TIMES	TCB TIME	PREEMPT SRB	NONPREEMPT SRB	TOTAL TIME	PREEMPT IIP SRB
SYSTEM SERVICES ADDRESS SPACE	1:39.995961	0.000000	3:25.079924	5:05.075886	N/A
DATABASE SERVICES ADDRESS SPACE	1:31.822012	0.000000	12:28:38.995808	12:30:10.817820	0.000000
IRLM	0.456105	0.000000	3:02.893287	3:03.349391	N/A
DDF ADDRESS SPACE	2.730084	20:28:36.142998	30:35.615420	20:59:14.488502	19:33:32.868978
TOTAL	3:15.004163	20:28:36.142998	13:05:42.584438	1 09:37:33.7316	19:33:32.868978

DB2 VERSION: V10 SCOPE: MEMBER TO: 11/11

---- HIGHLIGHTS ----

INTERVAL START	: 11/10/11 06:09:00.00	SAMPLING START:	11/10/11 06:09:00.00	TOTAL THREADS	: 290.00
INTERVAL END	: 11/11/11 06:06:00.00	SAMPLING END	: 11/11/11 06:06:00.00	TOTAL COMMITS	: 10749.2K
INTERVAL ELAPSED:	23:57:00.000072	OUTAGE ELAPSED:	0.000000	DATA SHARING MEMBER:	N/A

CPU TIMES	TCB TIME	PREEMPT SRB	NONPREEMPT SRB	TOTAL TIME	PREEMPT IIP SRB
SYSTEM SERVICES ADDRESS SPACE	2:26.595613	2:14.698997	13.547515	4:54.842125	N/A
DATABASE SERVICES ADDRESS SPACE	1:04.360185	5:49:17.448125	11.274434	5:50:33.082744	4:25:03.509555
IRLM	0.032864	0.000000	3:39.871402	3:39.904266	N/A
DDF ADDRESS SPACE	6.096981	2 22:30:18.7722	56:23.794572	2 23:26:48.6638	1 11:39:09.8193
TOTAL	3:37.085643	3 04:21:50.9193	1:00:28.487923	3 05:25:56.4929	1 16:04:13.3288

Parallelism: II14441, II12836

- V8
  - Only Serial tasks cost out by optimizer
  - Parallelism cut on first table
  - limited 1x processors
- V9
  - Optimizer costs parallel tasks
  - Parallelism can be cut on inner table
  - Limited by 4x processors

- V9 Utilities- PK41899 (SORTNUM)
  - Load, Reorg, Rebuild, CHECK = 3x CPUs
  - Unload = 1x CPUs
  - Parallel index Load, Reorg, Rebuild = unlimited

	Is parallelism allowed?		
If query uses this...	**I/O**	**CP**	**Sysplex**
Access via RID list (list prefetch and multiple index access)	Yes	Yes	No
Queries that return LOB values	Yes	Yes	No
Merge scan join on more than one column	Yes	Yes	Yes
Queries that qualify for direct row access	No	No	No
Materialized views or materialized nested table expressions at reference time	No	No	No
EXISTS within WHERE predicate	No	No	No
Security label column on table	Yes	Yes	No

Parallelism in DB2 10

- Parallelism is now enabled when a query involves a CTE reference, a table function, or a CREATE GLOBAL TEMPORARY table (CGTT), or a work file resulted from view materialization, table expression materialization, or full outer join.
- Parallelism is allowed when the optimizer chooses index reverse scan for a table.
- Parallelism is now enabled for multi-row fetch.
- Parallelism is allowed when the leading table is sort output and the join between the leading table and the second table is multiple column hybrid join.
- When parallelism is enabled, the optimizer can choose a hybrid join.
- The optimizer can perform subquery transformed to join when parallelism is enabled.
- http://publib.boulder.ibm.com/infocenter/dzichelp/v2r2/topic/com.ibm.db2z10.doc.perf/src/tpc/db2z_whenparallelnotused.htm

Parallelism in DB2 10…

- Previous releases of DB2 divide the number of keys or pages by the number representing the parallel degree
  - One task is allocated per degree of parallelism
  - The range is processed and the task ends
  - Tasks may take different times to process due to uneven distribution/skew

- DB2 10 may use the Straw Model workload distribution method
  - More key or page ranges will be allocated than the number of parallel degrees
  - The same number of tasks as before
  - Once a task finishes its smaller range it will process another range
  - Skewed data has the opportunity to be divided into a smaller number of pieces

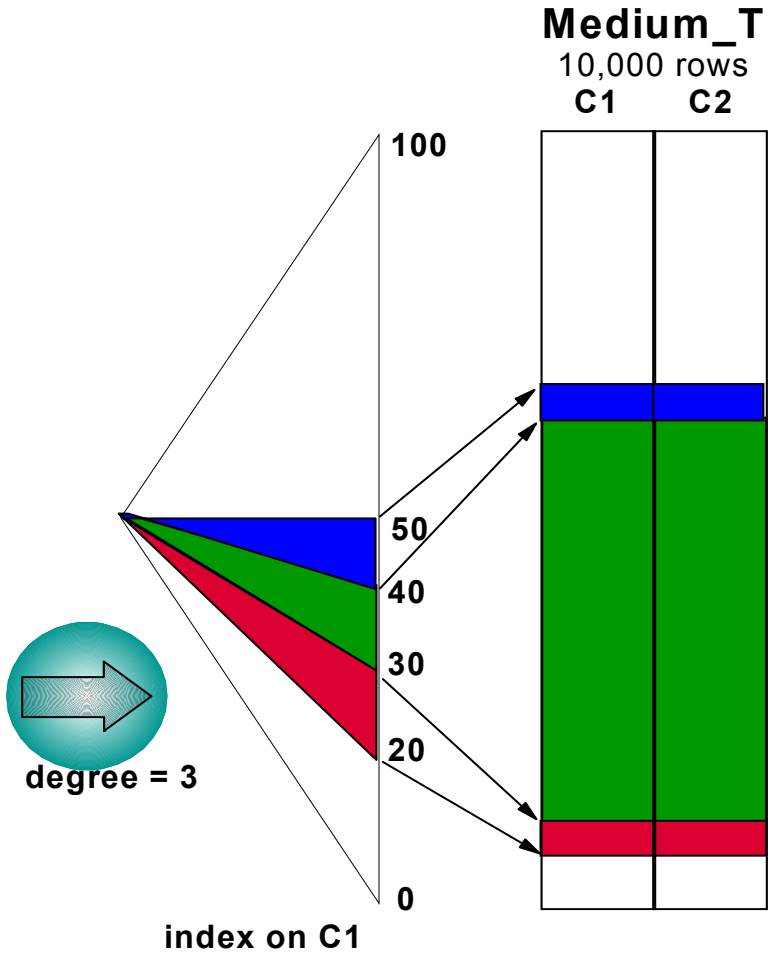
Parallelism... STRAW Model

```
SELECT *
FROM   Medium_T M
WHERE  M.C1 BETWEEN 20 AND 50
```

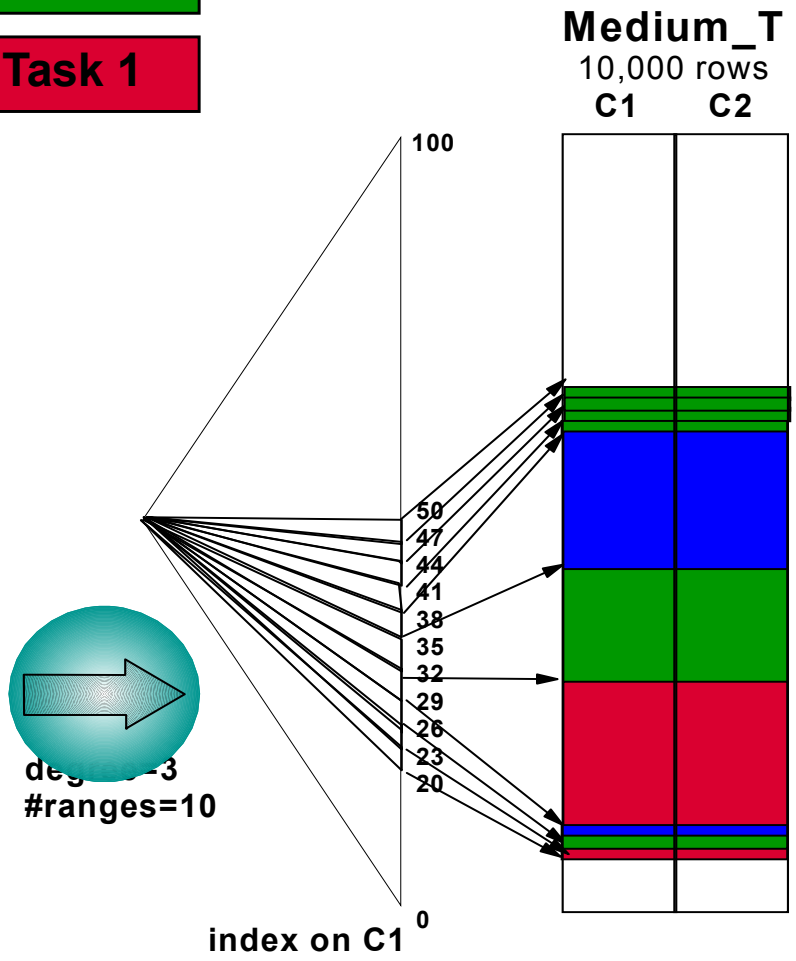
Task 3

Task 2

Task 1



Divided in key ranges before DB2 10



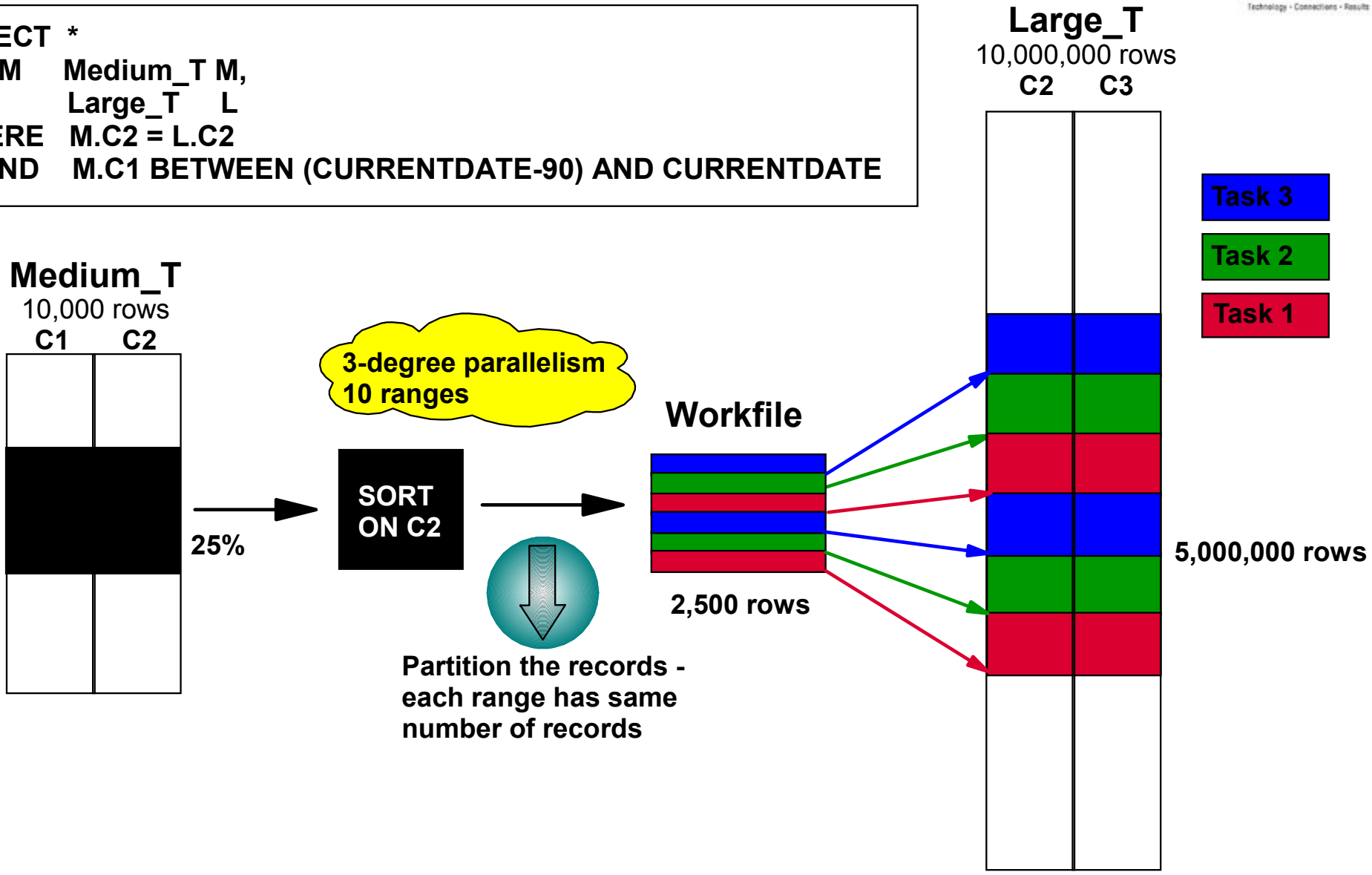
Divided in key ranges with Straw Model

Parallelism in DB2 10…

- Dynamic record range partitioning
  - Intermediate results are divided into ranges
    - Equal number of records
    - Division doesn't have to be on the key boundary
      - *Unless required for group by or distinct function*
  - Record range partitioning is dynamic
    - No longer based on the key ranges decided at bind time
  - Not impacted by
    - Data skew,
    - Out of date statistics
  - Now based on number of
    - Composite side records and
    - Workload elements
  - Will attempt to use in-memory work file for the materialization

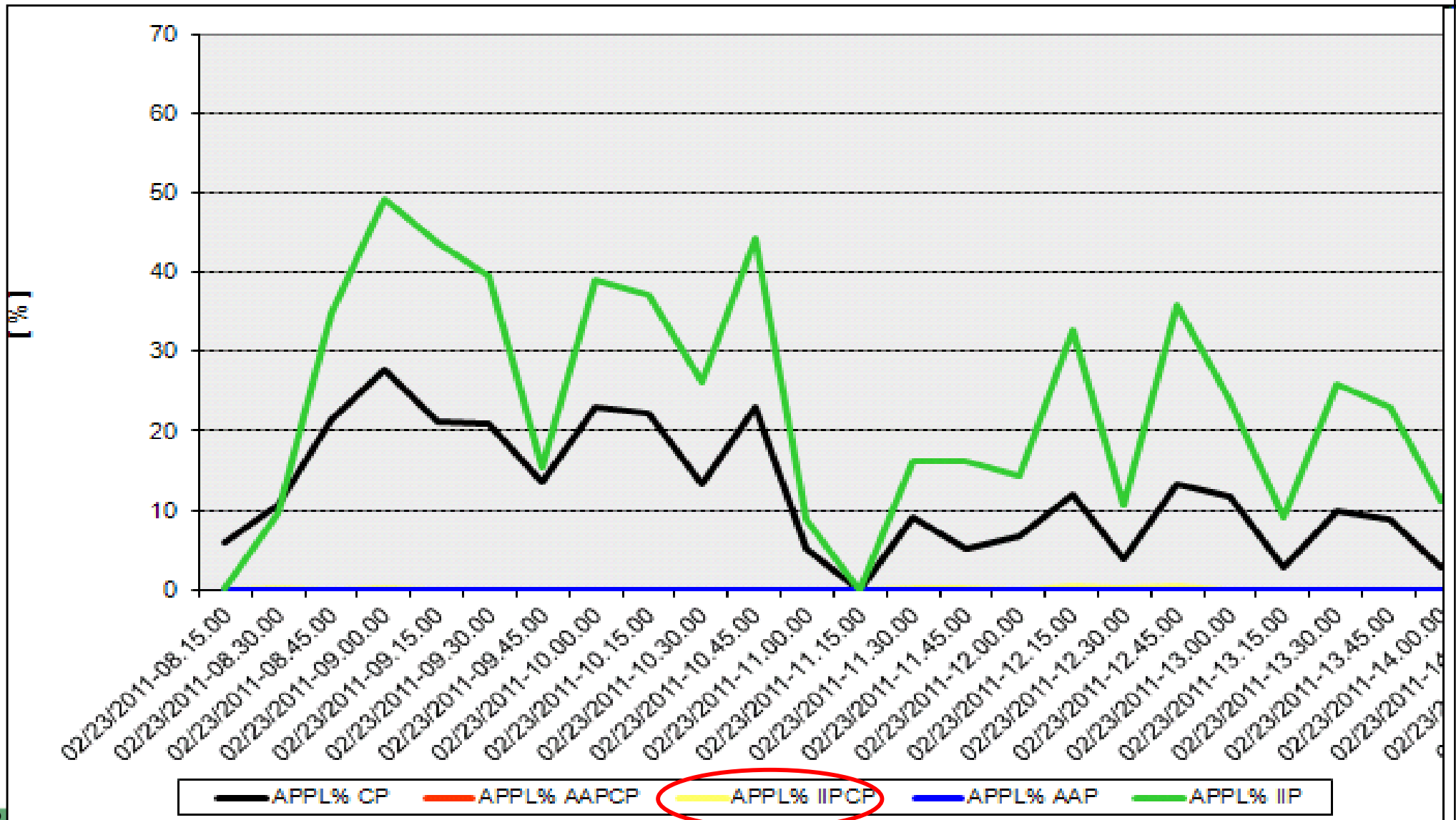
Parallelism...Dynamic record range partition

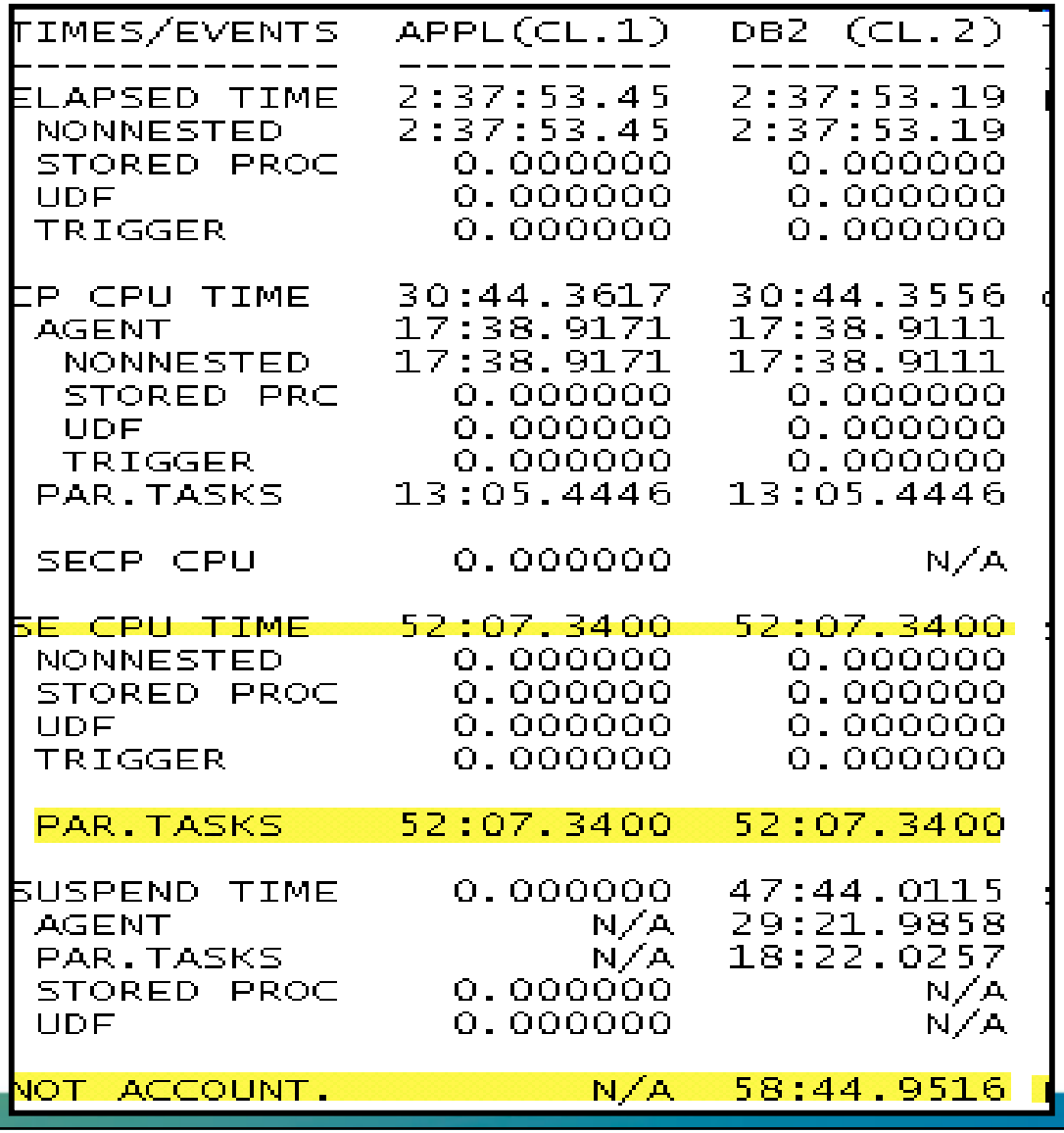
```
SELECT  *
FROM    Medium_T M,
        Large_T   L
WHERE   M.C2 = L.C2
  AND   M.C1 BETWEEN (CURRENTDATE-90) AND CURRENTDATE
```



Parallelism in production (case study)

- Remember a high % of parallel tasks becomes zIIP eligible



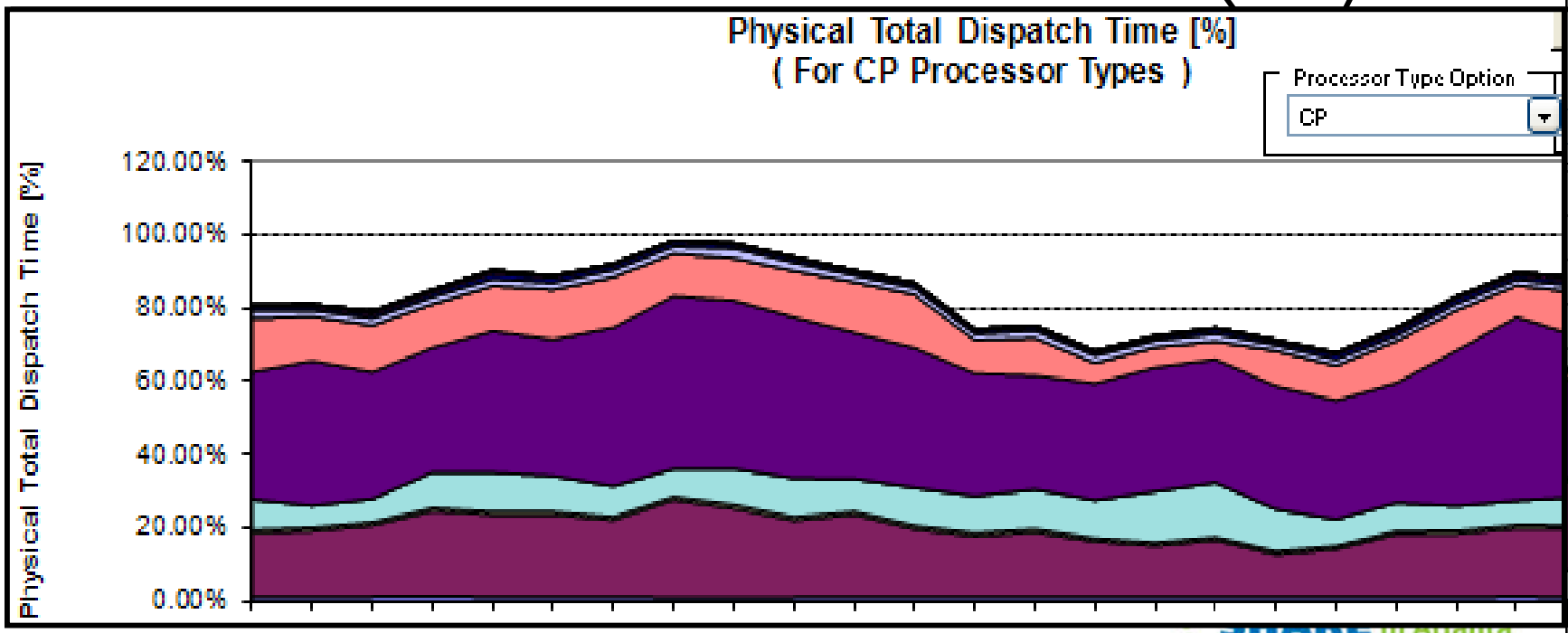


TIMES/EVENTS	APPL(CL.1)	DB2 (CL.2)
ELAPSED TIME	2:37:53.45	2:37:53.19
NONNESTED	2:37:53.45	2:37:53.19
STORED PROC	0.000000	0.000000
UDF	0.000000	0.000000
TRIGGER	0.000000	0.000000
CP CPU TIME	30:44.3617	30:44.3556
AGENT	17:38.9171	17:38.9111
NONNESTED	17:38.9171	17:38.9111
STORED PRC	0.000000	0.000000
UDF	0.000000	0.000000
TRIGGER	0.000000	0.000000
PAR.TASKS	13:05.4446	13:05.4446
SECP CPU	0.000000	N/A
SE CPU TIME	52:07.3400	52:07.3400
NONNESTED	0.000000	0.000000
STORED PROC	0.000000	0.000000
UDF	0.000000	0.000000
TRIGGER	0.000000	0.000000
PAR.TASKS	52:07.3400	52:07.3400
SUSPEND TIME	0.000000	47:44.0115
AGENT	N/A	29:21.9858
PAR.TASKS	N/A	18:22.0257
STORED PROC	0.000000	N/A
UDF	0.000000	N/A
NOT ACCOUNT.	N/A	58:44.9516

A peak at the LPAR

- LPAR trend report
  - The entire box is between 90%-100% at the end of the run
  - The zIIP is running over 80% at the time of the query
    - 4 CPs and 1 zIIP
    - Law of probability (each CP 40% utilized)
  - Easy to see constraint as no workload is making its goal

warehouse query and DB2



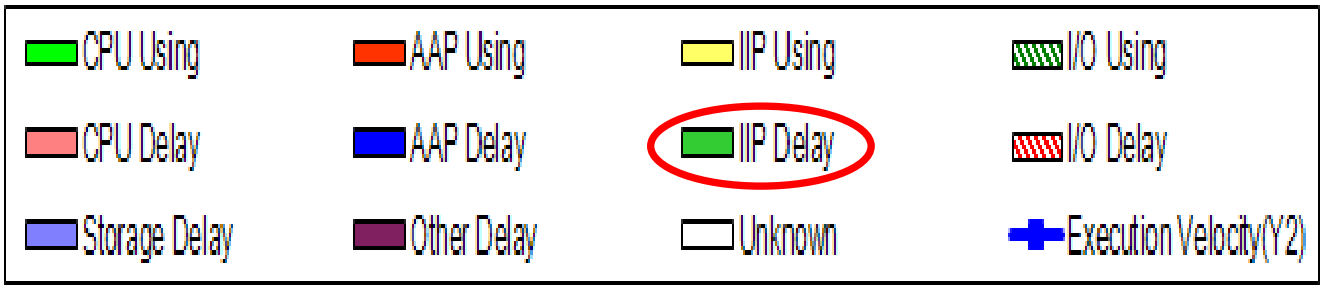
Parallelism investigation

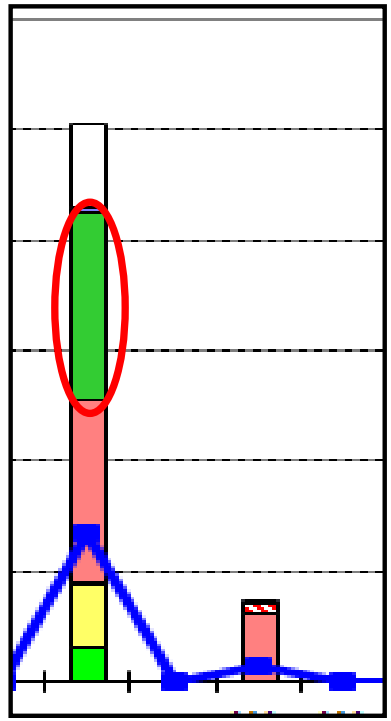
- RMF Spreadsheet Reporter Response delay report
  - Part of WLM Activity Report
- DSNTIJUZ
  - PARAMDEG?
  - BUFFERPOOL VPPSEQT??
- SYS1.PARMLIB (IEAOPTxx)
  - IIPHONORPRIORITY = **NO**
- Lots of unaccounted for time
  - OMPE accounting

QUERY PARALLEL.	TOTAL
MAXIMUM MEMBERS	N/P
MAXIMUM DEGREE	4
GROUPS EXECUTED	78
RAN AS PLANNED	78
RAN REDUCED	0
ONE DB2 COOR=N	0
ONE DB2 ISOLAT	0
ONE DB2 DCL TTABLE	0
SEQ - CURSOR	0
SEQ - NO ESA	0
SEQ - NO BUF	0
SEQ - ENCL.SER	0

CLASS 2 TIME DISTRIBUTION

CPU	=====> 11%
SECPU	
NOTACC	==========> 37%
SUSP	=========> 19%





CPU delay at about 33%, and the zIIP suspense time at 34%.

What to look for with parallelism

```
DSNB440I  DB1S PARALLEL ACTIVITY -
          PARALLEL REQUEST =      7  DEGRADED PARALLEL=      0
```

- DSNB440I - shows degraded parallel tasks from buffer pools
- DSNU397I – Utility message on constrained tasks (SORTNUM)
- -DISPLAY THREAD(*) – PT appears next to parallel tasks
- STATS long report – calculate BP size based on number of denied parallel tasks
- ACCNT trace – Query parallelism section
  - Ran as Planned/Ran reduced
- IFCID 0222 – OMEGAMON activity trace
  - Shows actual number of tasks and degradation
- IFCID 0221 – tells you which buffer pool restricted parallelism

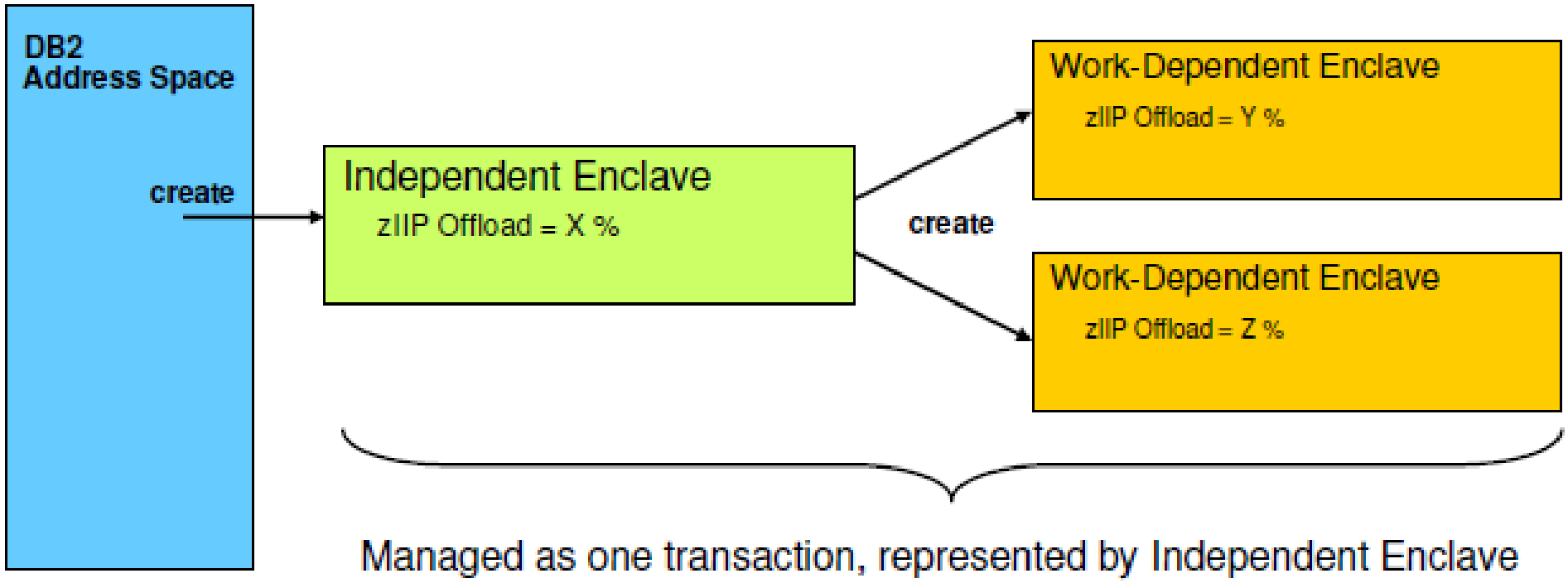
QUERY PARALLELISM	QUANTITY
MAX.DEGREE OF PARALLELISM	4.00
PARALLEL GROUPS EXECUTED	5.00
RAN AS PLANNED	5.00
RAN REDUCED	0.00
SEQUENTIAL-CURSOR	0.00
SEQUENTIAL-NO ESA	0.00
SEQUENTIAL-NO BUFFER	0.00
SEQUENTIAL-ENCLAVE SER.	0.00
ONE DB2 - COORDINATOR = NO	0.00
ONE DB2 - ISOLATION LEVEL	0.00
ONE DB2 - DCL TTABLE	0.00
MEMBER SKIPPED (%)	N/C
REFORM PARAL-CONFIG CHANGED	0.00
REFORM PARAL-NO BUFFER	0.00

What You Control for parallelism..

- Hidden zParm SPRMPTH – DSN6SPRC
  - Threshold below which parallelism disabled
- PARAMDEG – MAX_DEGREE limits parallel groups
  - Static and dynamic SQL (default '0', unlimited)
- ASSIST, COORDNTR – DS group parallelism (X type)
- DEGREE(ANY) and CURRENTDATA(NO) bind options
  - Or DB2 needs to know if cursor is readonly
- CDSSRDEF – SET CURRENT DEGREE special register for dynamic queries
  - Default =1, 'ANY' lets DB2 decide
- VPPSEQT - % of sequential steal for parallel operations
  - Each utility task needs 128 pages in BP
- PK41899 – (DB2 9) DFSORT removes need for SORTNUM in Utility statements, uses RTS or STATS if needed to calculate sizes (UTSORTAL)
- Star join enabled, number of tables involved
- PARA_EFF - **% of optimism regarding parallel access path improvement (PM16020)**

AccessPath	sequential_cost	parallel_degree	parallel_reduced_cost
AP1	1000	5	400
AP2	2000	20	300

z/OS 1.11 and 1.12



- Previously offload had to be consistent across and Enclave
- Work-dependent enclaves allow different zIIP offloads
  - WLM APAR OA26104 (z/OS 1.10)
  - DB2 APAR PK76676

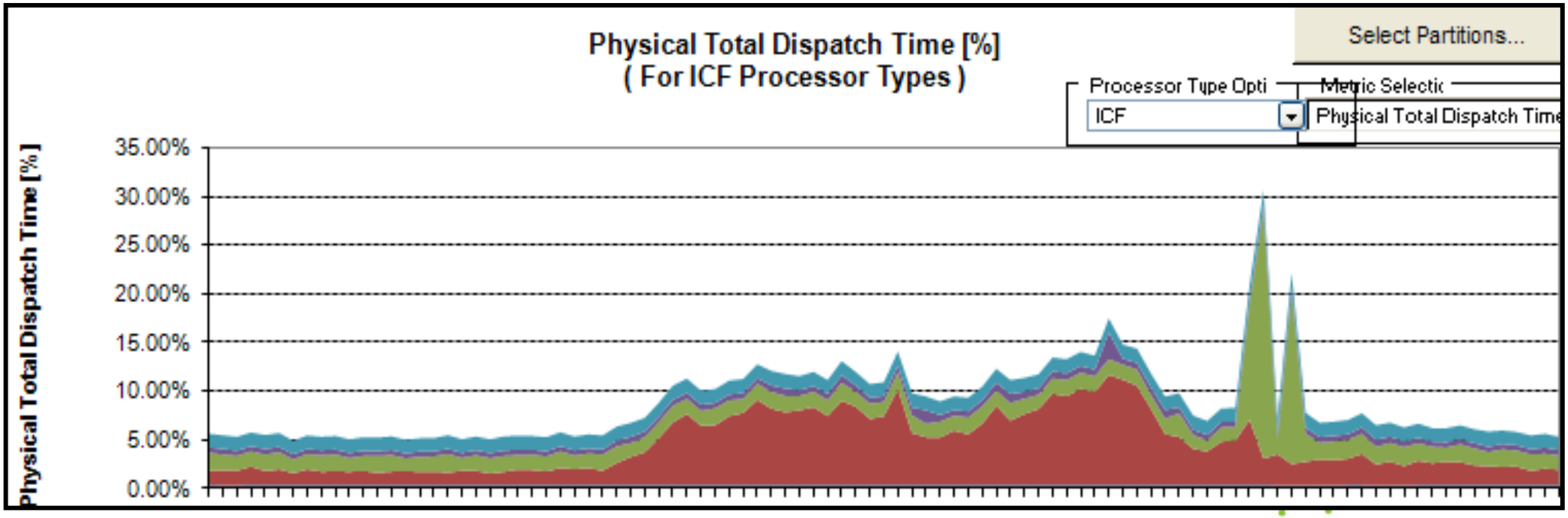
PARMLIB Parameters

- **ZIIPAWMT, ZAAPAWMT – Specify zIIP alternate wait management threshold**
  - Wake up time confused with busy time
  - Default 12 milliseconds
- **zAAP has other settings not applicable to zIIP**
  - IFACrossover – disallow zAAP work on general CP
- **ZAAPZIIP = YES|NO (IEAOPTxx option)**
  - **Allows zAAP eligible workload to run on a zIIP**
  - **If HONORPRIORITY=YES then you should enable Alternate wait management**
  - APAR OA20045 added IIPHONORPRIORITY to IEAOPT parmlib member

** Be careful about attempting to FORCE zIIP offload

zAAP notes

- zAAP (debuted in 2004)
  - on old hardware like z900, z990 zAAP time can fall under ICF (also CMF can report it incorrectly)
    - look under Partition data in (CPU,LPAR, Paging) report and see if there are duplicate names of LPARs



Offload APARS

- PM12256 – zIIP offload improvement up to 60%, and less overhead
  - http://www-01.ibm.com/support/docview.wss?uid=swg1PM12256&myns=swgimgmt&mynp=OCSSEPEK&mync=R
- PM28626 – corrected PM12256
  - https://www-304.ibm.com/support/docview.wss?crawler=1&uid=swg1PM28626
- OA35146 – z/OS for PM12256
  - https://www-304.ibm.com/support/entdocview.wss?uid=isg1OA35146

Results of zIIP Maint.

- Pre- PM12256

```
CLASS 2 TIME DISTRIBUTION
-------------------------------------------------
CPU     |=========> 19%
SECPU   |=> 3%
NOTACC  |====> 9%
SUSP    |=================================> 70%
```

- After – PM12256

```
CLASS 2 TIME DISTRIBUTION
-------------------------------------
CPU     |
SECPU   |======> 12%
NOTACC  |======================> 43%
SUSP    |======================> 45%
```

- After – PM28626 ???
  - Less noticeable elapsed time difference for customers with knee-capped general CPs

Reference material

- II12836 – Info APAR for parallelism V6- DB2 9
- PK19920 – UK15814 maintenance for Utility processing
- II14219 – zIIP Exploitation
- PK27578 – zIIP for parallel queries
- PK18454 – zIIP use for DRDA threads
- OA37201 – faster switch to SRM mode
- OA38155 – avoid ABENDs due to z/OS changes
- PM06953 – parallel tasks under 1 enclave (08/10)
  - http://www-01.ibm.com/support/docview.wss?uid=swg1PM06953
- Techline Sizing with CP3000 tool: contact your local IBMer
- RMF Spreadsheet Reporting Tool
  - http://www-03.ibm.com/systems/z/os/zos/features/rmf/tools/rmftools.html
- Getting Started Resources
  - http://www-03.ibm.com/systems/z/hardware/features/ziip/resources.html

Questions???

Adrian Burke
DB2 SWAT Team SVL Lab
agburke@us.ibm.com

- VISIT the DB2 Best Practices
- VISIT the DB2 for z/OS Exchange
- JOIN the World of DB2 for z/OS
- JOIN the DB2 for z/OS group

zIIP Software Enablement Process

- Install z/OS zIIP support maintenance (II14219)
- Install DB2 for z/OS support maintenance
  - http://www-03.ibm.com/systems/z/os/zos/downloads/
- Set up SYS1.PARMLIB(IEAOPTxx) member
  - **When zIIP hardware is not installed set PROJECTCPU=YES for projecting zIIP redirect**
  - **zIIP redirect projection / estimation is shown under APPL% IIPCP in the RMF Workload Activity Report and under IIPCP CPU in the IBM Tivoli Omegamon DB2PE Accounting Report**