DB2 for z/OS and zIIP & zAAP

Timm Zimmermann & Adrian Burke
IBM DB2 for z/OS Development

tizimm@de.ibm.com
aburke@us.ibm.com
Important Disclaimers

Disclaimer:
Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The Information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance Disclaimer:
This document contains performance information based on measurements done in a controlled environment. The actual throughput or performance 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 or performance improvements equivalent to the numbers stated here.
Agenda

• Specialty Engines
• How can SE help me
• What is eligible as of V10?
• Where do I look?
  – Customer example
• What can I control?
  – Parallelism
• Recent enhancements
Mainframe Innovation: Specialty Engines

Eligible for zIIP:
- DB2 remote access, BI/DW, Utilities Build Index and Sort processing, XML Parsing, RUNSTATS, BP Prefetch, Deferred Write
- z/OS XML System Services
- HiperSockets for large messages
- IPSec encryption
- z/OS Global Mirror (XRC)
- IBM GBS Scalable Architecture for Financial Reporting
- z/OS CIM Server
- ISVs

Eligible for zAAP:
- Java execution environment
- z/OS XML System Services
The IBM System z specialty engines can operate on the same machine together.
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: MSU units, generally increases with the # of general processors
- BUT/AND…. it can also result in latent demand processing so processor utilization remains constant
Work is dispatched

• Standard dispatchable units (DUs) are the Task Control Block and the Service Request Block
  – TCB runs at dispatching priority of address space and is pre-emptible
  – SRB runs at supervisory priority and is non-pre-emptible

• Advanced dispatching units
  – Enclave
    • Anchor for an address space-independent transaction managed by WLM
    • Can comprise multiple DUs (TCBs and Enclave SRBs) executing across multiple address spaces
  – Client SRB
    • Created and executed like an ordinary SRB but runs with (Scheduler) dispatching priority and is pre-emptible
  – Enclave SRB
    • Created and executed like an ordinary SRB but runs with Enclave dispatching priority and is pre-emptible
Work is dispatched

• 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.

• Pre-emptible enclaves are used to do the work on behalf of the originating address space. Enclaves are grouped by common characteristics into service classes and since they are pre-emptible, the z/OS dispatcher can interrupt these tasks for more important ones. There are two types of pre-emptible SRBs: client SRBs and enclave SRBs.

• If the DB2 for z/OS request is coming in over distributed (i.e., 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.
## TCB and SRB Times – Major Contributors in DB2

<table>
<thead>
<tr>
<th>TCB</th>
<th>Application</th>
<th>Synch I/O</th>
<th>Logical logging</th>
<th>GBP reads*</th>
<th>Dataset Open/Close</th>
<th>DBM1 Full System Contraction</th>
<th>Preformat</th>
<th>Extend</th>
<th>DBAS</th>
<th>SQL processing</th>
<th>Lock requests</th>
<th>Logical logging</th>
<th>Buffer updates</th>
<th>GBP reads*</th>
<th>Global lock requests*</th>
<th>SCMS</th>
<th>Prefetch read</th>
<th>Deferred write</th>
<th>Parallel child tasks</th>
<th>Physical log write</th>
<th>Thread deallocation</th>
<th>Update commit incl. page P-lock unlock*</th>
<th>Backouts</th>
<th>Checkpoints</th>
<th>Error checking</th>
<th>Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRB</td>
<td>Application</td>
<td>Synch I/O</td>
<td>Logical logging</td>
<td>GBP reads*</td>
<td>Dataset Open/Close</td>
<td>DBM1 Full System Contraction</td>
<td>Preformat</td>
<td>Extend</td>
<td>DBAS</td>
<td>SQL processing</td>
<td>Lock requests</td>
<td>Logical logging</td>
<td>Buffer updates</td>
<td>GBP reads*</td>
<td>Global lock requests*</td>
<td>SCMS</td>
<td>Prefetch read</td>
<td>Deferred write</td>
<td>Parallel child tasks</td>
<td>Physical log write</td>
<td>Thread deallocation</td>
<td>Update commit incl. page P-lock unlock*</td>
<td>Backouts</td>
<td>Checkpoints</td>
<td>Error checking</td>
<td>Management</td>
</tr>
</tbody>
</table>

*Data Sharing specific

(*) Eligible to run on 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

• The Open Group adopted DRDA in 1998 as the open standard for database access interoperability (DB2, Informix, Oracle)
  • Private Protocol, also IBM, was stabilized in V4 of DB2
  • DRDA can use TCP/IP or SNA as a network protocol to flow commands

How does that affect 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 due to double buffering (private shared)
⇒ Private Protocol support is removed in V10
How many zIIPs should I have?

- 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
  - Remember to normalize

http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/PRS3948
**zIIP usage**

- How many zIIPs do you need (this scenario 12:1 ratio CP to zIIP)
  - zIIP eligible work that went to CP either zIIP is overloaded, or local suspense lock - Red line on graph (AAPL% IIPCP)
  - Must have enough capacity to absorb spikes, not just typical offload
  - Needs Help algorithm ensures work does not pile up waiting on zIIP
- **Law of probability for many CPs vs. zIIPs** (next slide)
zIIP Overflow

- If 12 CPs are 65% (0.65^12) utilized then each CP is 0.5% instantaneously busy
  - If 1 zIIP is 35% busy it is instantaneously 35% busy
  - So with ‘needs help’ algorithm it is likely some zIIP eligible work could fall back to a CP
    - See IIPHONORPRIORITY slide 31
- Markov’s Equation is based on 1 server (CP) in steady state
- As Utilization approaches 100% wait time approaches ∞
  - this will cause more work to overflow to a CP starting at around 35% utilization of a single zIIP processor
  - More zIIPs = more offload

Tw = wait time of transaction
Ts = service time of transaction
U = utilization
The knee of the curve occurs at 35% for 1 processor, thereafter Tw increases drastically

Tw = Ts * U / (1-U)
What is new with zIIP & zAAP OMPE reporting?

• Tivoli Omegamon DB2PE Accounting report changes with APARs PK51045 & PK50575:
  – IIP changed to SE to indicate that the value may include CPU usage from either or both Specialty Engines (zIIP & zAAP)
  – Example:
    • SE CP CPU will include zIIP CPU and zAAP CPU for a DRDA SQLJ Java External Stored Procedure workload

• SECP (projection / overflow) does not include zAAP overflow or zAAP projection
  – Applicable only to zIIP in DB2 V8 and DB2 9

• SECP (projection / overflow) is not reported in DB2 10.
  – Need to use RMF Workload Activity Report Service / Reporting Class information
What is new with DB2 address space CPU reporting?

- OMPE Statistics Report CPU section shows the zIIP redirect for DB2 10 DBM1 Prefetch and Deferred Write processing (with APAR PM30468)

<table>
<thead>
<tr>
<th>CPU TIMES</th>
<th>TCB TIME</th>
<th>PREEMPT SRB</th>
<th>NONPREEMPT SRB</th>
<th>TOTAL TIME</th>
<th>PREEMPT IIP SRB</th>
<th>/COMMIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>SYSTEM SERVICES ADDRESS SPACE</td>
<td>0.012387</td>
<td>0.000777</td>
<td>0.003326</td>
<td>0.016490</td>
<td>N/A</td>
<td>0.008245</td>
</tr>
<tr>
<td>DATABASE SERVICES ADDRESS SPACE</td>
<td>0.375790</td>
<td>24.036392</td>
<td>0.002450</td>
<td>24.414632</td>
<td>1:03.304060</td>
<td>12.207316</td>
</tr>
<tr>
<td>IRLM</td>
<td>0.000018</td>
<td>0.000000</td>
<td>0.078741</td>
<td>0.078759</td>
<td>N/A</td>
<td>0.039379</td>
</tr>
<tr>
<td>DDF ADDRESS SPACE</td>
<td>0.000267</td>
<td>0.000018</td>
<td>0.000133</td>
<td>0.000418</td>
<td>0.000000</td>
<td>0.000209</td>
</tr>
<tr>
<td>TOTAL</td>
<td>0.388461</td>
<td>24.037188</td>
<td>0.084651</td>
<td>24.510300</td>
<td>1:03.304060</td>
<td>12.255150</td>
</tr>
</tbody>
</table>

- PREEMPT IIP SRB time shows the CPU time redirected to zIIP
- /COMMIT shows the chargeable (non-zIIP) CP CPU time
What does it look like in an accounting report

<table>
<thead>
<tr>
<th></th>
<th>AVERAGE</th>
<th>APPL (CL.1)</th>
<th>DB2 (CL.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP CPU TIME</td>
<td>19.373768</td>
<td>19.365788</td>
<td></td>
</tr>
<tr>
<td>AGENT</td>
<td>6.779348</td>
<td>6.771411</td>
<td></td>
</tr>
<tr>
<td>NONNESTED</td>
<td>6.779348</td>
<td>6.771411</td>
<td></td>
</tr>
<tr>
<td>STORED PRC</td>
<td>0.000000</td>
<td>0.000000</td>
<td></td>
</tr>
<tr>
<td>UDF</td>
<td>0.000000</td>
<td>0.000000</td>
<td></td>
</tr>
<tr>
<td>TRIGGER</td>
<td>0.000000</td>
<td>0.000000</td>
<td></td>
</tr>
<tr>
<td>PAR.TASKS</td>
<td>12.594420</td>
<td>12.594377</td>
<td></td>
</tr>
<tr>
<td>SECP CPU</td>
<td>2.813831</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>SE CPU TIME</td>
<td>35.886951</td>
<td>35.886951</td>
<td></td>
</tr>
</tbody>
</table>

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% ((SE + SECP) / (CP + SE))

zIIP Redirect % = 65% ((SE / (CP + SE))

zIIP eligible but ran on CP = 5% ((SECP / (CP + SE))
### What is Eligible?

<table>
<thead>
<tr>
<th>zIIP Eligible</th>
<th>Function</th>
<th>Amount Redirected</th>
<th>Prerequisites</th>
</tr>
</thead>
</table>
| **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 (depending on # of parts and indexes (BUILD and REBUILD phases of index maintenance go to zIIP)  
                                             2) Up to 55-60% in Lab measurements.  
                                             3) Portion of main task for remote calls, 80% of child tasks  
                                             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 |
## What is Eligible? …

<table>
<thead>
<tr>
<th><strong>zIIP Eligible</strong></th>
<th><strong>Function</strong></th>
<th><strong>Amount Redirected</strong></th>
<th><strong>Prerequisites</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>APAR II14219</strong></td>
<td><strong>1)</strong> All of DB2 v8 and 9 offload++</td>
<td><strong>1)</strong> BUILD phase, Native SQL procs, parallelism, 60% DRDA requests</td>
<td><strong>1)</strong> DB2 10 / z/OS 1.10</td>
</tr>
<tr>
<td></td>
<td><strong>2)</strong> RUNSTATS</td>
<td><strong>2)</strong> Basic RUNSTATS for table, NO Histogram, DSTATS, COLGROUP … BUT index stats almost all offloaded (not DPSIs)</td>
<td><strong>2)</strong> Run RUNSTATS, no inline STATS</td>
</tr>
<tr>
<td></td>
<td><strong>3)</strong> Prefetch and deferred write processing</td>
<td><strong>3)</strong> 100% (roughly 70% of DBM1 SRB time)</td>
<td><strong>3)</strong> Shows up in DBM1 SRB time</td>
</tr>
<tr>
<td></td>
<td><strong>4)</strong> Parallelism enhancements</td>
<td><strong>4)</strong> Parallelism more likely (80% of child tasks)</td>
<td><strong>4)</strong> V10 NFM with rebinding</td>
</tr>
<tr>
<td><strong>DB2 10</strong></td>
<td><strong>1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DB2 11?</strong></td>
<td><strong>1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>1)</strong> IPSec</td>
<td><strong>1)</strong> Encryption processing, header processing and crypto validation (93% for bulk data movement)</td>
<td><strong>1)</strong> z/OS 1.8 + UA34582 AND z/OS Communication Server PTF UK27062-63</td>
</tr>
<tr>
<td></td>
<td><strong>2)</strong> Global Mirror for z/OS (formerly Extended Remote Copy)</td>
<td><strong>2)</strong> Most System Data Mover processing</td>
<td><strong>2)</strong> z/OS 1.10, or 1.9 + UA39510, or 1.8 + UA39509 (zGM parmlib zIIPEnable)</td>
</tr>
<tr>
<td></td>
<td><strong>3)</strong> HiperSockets for Large messages</td>
<td><strong>3)</strong> Handles large outbound messages (multiple channel paths given to SRBs)</td>
<td><strong>3)</strong> z10 and z/OS 1.10 (GLOBALCONFIG ZIIP IQDIOMULTIWRITE)</td>
</tr>
<tr>
<td></td>
<td><strong>4)</strong> DFSORT</td>
<td><strong>4)</strong> Sorting of fixed length rows (10-40% Utility), memory object work file sorts</td>
<td><strong>4)</strong> PK85899 and PK85856 (z/OS 1.10), PM62824 and z/OS 1.12</td>
</tr>
<tr>
<td></td>
<td><strong>5)</strong> zAAP on zIIP</td>
<td><strong>5)</strong> zAAP eligible work can move to zIIP if no zAAP installed</td>
<td><strong>5)</strong> z/OS 1.11 base + z/OS 1.10 OR APAR OA27495 / OA38829 if both installed</td>
</tr>
</tbody>
</table>

---

### Other Processes

| **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 (10-40% Utility), memory object work file sorts | **5)** zAAP eligible work can move to zIIP if no zAAP installed |

---

*Read left to right all the way across 1) Function -> 1) Amount Redirected -> 1) Prerequisites*
zAAP on zIIP Capability

• A new capability that can enable System z Application Assist Processor (zAAP) eligible workloads to run on System z Integrated Information Processors (zIIPs).
  – For customers with no zAAPs and zIIPs
    • The combined eligible workloads may make the acquisition of a single zIIP cost effective.
  – For customers with only zIIP processors
    • Makes Java and z/OS XML System Services -based workloads eligible to run on existing zIIPs – maximizes zIIP investment.
  – Available on z/OS V1.9, V1.10 and V1.11
    • This new capability is not available for z/OS LPARS if zAAPs are installed on the server.

• zAAP on zIIP planning guide -
  http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD103548
How to enable the zAAP on zIIP Capability

• The capability ships default enabled with z/OS V1.11.
  – Parameter in SYS1.PARMLIB(IEASYSxx) : ZAAPZIIP = YES (default in z/OS V1.11)
  – If you wish to disable the function for any reason, you must IPL with ZAAPZIIP=NO in the IEASYSxx Parmlib member.
  – IFAHONORPRIORITY now based on IIPHONORPRIORITY

• Also available with z/OS V1.9 and V1.10
  – With PTF for APAR OA27495, and
  – Enabled with ZAAPZIIP=YES in the IEASYSxx Parmlib (the default is NO)

• This new capability does not remove the requirement to purchase and maintain one or more general purpose processors for every zIIP processor on the server.
  – Possible performance degradation if 1CP:1zIIP:1zAAP before

• APAR OA38829 for z/OS 1.12 and 1.13
  – Allow zAAP on zIIP even if there is a zAAP installed
What is eligible – even more …

• Measured LOAD, REBUILD INDEX and REORG Utilities.
• zIIP redirect % depends on % CPU consumed by the Build Index phase of the Utility.
• Observed Class 1 CPU reduction for configuration with 4 CPs and 2 zIIPs with fixed length Index key:
  – 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
• CPU overhead incurred during execution unit switch from TCB to enclave SRB during Index Rebuild phase
  – Typically less than 10%
  – Eligible for zIIP redirect
Utility zIIP redirect with DFSORT

• Introduced in Aug 2009 for zIIP redirect for DFSORT processing for some DB2 Utilities
  – Applicable to in-memory fixed length record sort processing in DFSORT
• Utilities that benefit:
  – LOAD, REORG, REBUILD INDEX and CHECK INDEX for Index key Sort processing
  – CHECK DATA for Foreign key Sort processing
  – RUNSTATS for COLGROUP processing
• Measured zIIP redirect benefit
  – 30% to 60% of DFSORT CPU
  – 10% to 40% of total Utility CPU
  – Varies with number of Indices
    • More benefit with more Indices
    • Measurement with up to 6 Indices
• DFSORT with MSGICE256I DFSORT CODE IS ELIGIBLE TO USE ZIIP FOR THIS DB2 UTILITY RUN (PK85899)
Thread footwork T2 vs. T4 Connections same LPAR

- Going through DIST adds network translations, another address space, and context switch to an SRB – but you get zIIP offload
  - DB2 objective is to improve T2 driver performance to beat T4
  - Even local T4 connection hits DIST and TCP stack – need DDF WLM service class
  - Moving T2 to T4 is not 1 for 1 MIP exchange due to overhead
    - But, the worse behaving the application the more there is to offload so in the end it must be tested and compared
zIIP vs. GP while changing to T4 driver

- In 1 day customer switched all T2 to T4 connections, WPS and DB2 z/OS both on same LPAR… Not A Benchmark
  - GP MIPS from 350 → 230
  - zIIP MIPS from about 100 → 405 ….. Not a 1 for 1 exchange
# Stored Procedures with zIIPs

<table>
<thead>
<tr>
<th>Language</th>
<th>Base Billable Cost</th>
<th>Billable Cost after zIIP and/or zAAP acceleration</th>
</tr>
</thead>
<tbody>
<tr>
<td>COBOL stored proc</td>
<td>X (Baseline)</td>
<td>.88x</td>
</tr>
<tr>
<td>C stored proc</td>
<td>.95x</td>
<td>.83x</td>
</tr>
<tr>
<td>Remote SQLJ</td>
<td>1.78x</td>
<td>1.06x</td>
</tr>
<tr>
<td>SQLJ stored proc</td>
<td>1.21x</td>
<td>1.15x (zIIP + zAAP)</td>
</tr>
<tr>
<td>JDBC stored proc</td>
<td>2.11x</td>
<td>1.76x (zIIP + zAAP)</td>
</tr>
<tr>
<td>External SQL stored proc</td>
<td>1.62x</td>
<td>1.49x</td>
</tr>
<tr>
<td>Native SQL stored proc</td>
<td>1.14x</td>
<td>.65x</td>
</tr>
</tbody>
</table>
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
Parallelism has highest offload %: 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 operations = no limit

<table>
<thead>
<tr>
<th>If query uses this...</th>
<th>Is parallelism allowed?</th>
</tr>
</thead>
<tbody>
<tr>
<td>I/O</td>
<td>CP</td>
</tr>
<tr>
<td>Access via RID list (list prefetch and multiple index access)</td>
<td>Yes</td>
</tr>
<tr>
<td>Queries that return LOB values</td>
<td>Yes</td>
</tr>
<tr>
<td>Merge scan join on more than one column</td>
<td>Yes</td>
</tr>
<tr>
<td>Queries that qualify for direct row access</td>
<td>No</td>
</tr>
<tr>
<td>Materialized views or materialized nested table expressions at reference time</td>
<td>No</td>
</tr>
<tr>
<td>EXISTS within WHERE predicate</td>
<td>No</td>
</tr>
<tr>
<td>Security label column on table</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Parallelism in V10

- Lifted many previous limitations
  - Multi-row fetch
  - reverse index scan
  - full outer joins
  - CTEs
  - CGTTs
  - …

- Improves parallel processing
  - DB2 10 may use the Straw Model workload distribution method
    - More key or page ranges will be allocated than the number of parallel degrees
    - Once a task finishes its smaller range it will process another range
Parallelism in production (case study)

- Remember a 80% of parallel child tasks are zIIP eligible
  - Here we see there are no zIIP cycles that went to a GP

<table>
<thead>
<tr>
<th>TIMES/EVENTS</th>
<th>APPL (CL.1)</th>
<th>DB2 (CL.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELAPSED TIME</td>
<td>2:37:53.45</td>
<td>2:37:53.19</td>
</tr>
<tr>
<td>NONNESTED</td>
<td>2:37:53.45</td>
<td>2:37:53.19</td>
</tr>
<tr>
<td>STORED PROC</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>UDF</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>TRIGGER</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>CP CPU TIME</td>
<td>30:44.3617</td>
<td>30:44.3556</td>
</tr>
<tr>
<td>AGENT</td>
<td>17:38.9171</td>
<td>17:38.9111</td>
</tr>
<tr>
<td>NONNESTED</td>
<td>17:38.9171</td>
<td>17:38.9111</td>
</tr>
<tr>
<td>STORED PROC</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>UDF</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>TRIGGER</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>PAR.TASKS</td>
<td>13:05.4446</td>
<td>13:05.4446</td>
</tr>
<tr>
<td>SECP CPU</td>
<td>0.000000</td>
<td>N/A</td>
</tr>
<tr>
<td>EW CPU TIME</td>
<td>52:07.3400</td>
<td>52:07.3400</td>
</tr>
<tr>
<td>NONNESTED</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>STORED PROC</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>UDF</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>TRIGGER</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>PAR.TASKS</td>
<td>52:07.3400</td>
<td>52:07.3400</td>
</tr>
<tr>
<td>SUSPEND TIME</td>
<td>0.000000</td>
<td>47:44.0115</td>
</tr>
<tr>
<td>AGENT</td>
<td>N/A</td>
<td>29:21.9858</td>
</tr>
<tr>
<td>PAR.TASKS</td>
<td>N/A</td>
<td>18:22.0237</td>
</tr>
<tr>
<td>STORED PROC</td>
<td>0.000000</td>
<td>N/A</td>
</tr>
<tr>
<td>UDF</td>
<td>0.000000</td>
<td>N/A</td>
</tr>
<tr>
<td>NOT ACCOUNTABLE</td>
<td>N/A</td>
<td>58:44.9516</td>
</tr>
</tbody>
</table>
Parallelism Investigation (case study)

- RMF Spreadsheet Reporter Response delay report
  - Part of WLM Activity report
- Lots of unaccounted for time
  - OMPE accounting
  - This block does not show child task class 2 time

```
CLASS 2 TIME DISTRIBUTION

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>=-----</td>
</tr>
<tr>
<td>SECPU</td>
<td></td>
</tr>
<tr>
<td>NOTACC</td>
<td>===========</td>
</tr>
<tr>
<td>SUSP</td>
<td>======</td>
</tr>
</tbody>
</table>
```

- SYS1.PARMLIB (IEAOPTxx)
  - IIPHONORPRIORITY = NO
    - 3 parallel tasks waiting for 1 zIIP
  - ZIIPAWMT also important here if delays seen in I/O in V10
    - This becomes very important in V10 and V11

```
CPU delay at about 33%, and the zIIP suspense time at 34%.
```
What to look for with parallelism

• DSNDB440I – shows degraded parallel tasks from buffer pools
  – PARALLEL REQUEST and DEGRADED PARALLEL
• 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 and degradation
• IFCID 0221 – OMEGAMON activity trace
  – tells you which buffer pool restricted parallelism

<table>
<thead>
<tr>
<th>QUERY PARALLELISM</th>
<th>QUANTITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX, DEGREE OF PARALLELISM</td>
<td>4.00</td>
</tr>
<tr>
<td>PARALLEL GROUPS EXECUTED</td>
<td>5.00</td>
</tr>
<tr>
<td>RAN AS PLANNED</td>
<td>5.00</td>
</tr>
<tr>
<td>RAN REDUCED</td>
<td>0.00</td>
</tr>
<tr>
<td>SEQUENTIAL-CURSOR</td>
<td>0.00</td>
</tr>
<tr>
<td>SEQUENTIAL-NO ESA</td>
<td>0.00</td>
</tr>
<tr>
<td>SEQUENTIAL-NO BUFFER</td>
<td>0.00</td>
</tr>
<tr>
<td>SEQUENTIAL-ENCLAVE SER</td>
<td>0.00</td>
</tr>
<tr>
<td>ONE DB2 - COORDINATOR = NO</td>
<td>0.00</td>
</tr>
<tr>
<td>ONE DB2 - ISOLATION LEVEL</td>
<td>0.00</td>
</tr>
<tr>
<td>ONE DB2 - DCL TTABLE</td>
<td>0.00</td>
</tr>
<tr>
<td>MEMBER SKIPPED (%)</td>
<td>N/C</td>
</tr>
<tr>
<td>REFORM PARAL-CONFIG CHANGED</td>
<td>0.00</td>
</tr>
<tr>
<td>REFORM PARAL-NO BUFFER</td>
<td>0.00</td>
</tr>
</tbody>
</table>
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)
- DEGREE(ANY) and CURRENTDATA(NO) bind options
  - Or DB2 needs to know if cursor is read-only
- 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
- Star join enabled, number of tables involved

- PARA_EFF - % of optimism regarding parallel access path improvement (PM16020)

<table>
<thead>
<tr>
<th>AccessPath</th>
<th>sequential_cost</th>
<th>parallel_degree</th>
<th>parallel_reduced_cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP1</td>
<td>1000</td>
<td>5</td>
<td>400</td>
</tr>
<tr>
<td>AP2</td>
<td>2000</td>
<td>20</td>
<td>300</td>
</tr>
</tbody>
</table>
z/OS 1.11 and 1.12

- Previously offload was consistent across an enclave, needed an independent per child task – troublesome for RLF
- Work-dependent enclaves allow different zIIP offloads
  - WLM APAR OA26104 (z/OS 1.10)
  - DB2 APAR PK76676
PARMLIB Parameters affecting zIIP

• IIPHONORPRIORITY (YES|NO) in IEAOPTxx parmlib member
  – This means if we reach ZIIPMAXQL and ZIIPAWMT is triggered
    the dispatcher will route work over to a GP
• ZIIPMAXQL – Maximum number of dispatchable units that will queue waiting for a zIIP processor
  – Current default is 7

• ZIIPAWMT, ZAAPAWMT – Alternate wait management threshold is how long zIIP will run before checking to see if it needs help from GP
  – Default 12 milliseconds/ 3200 for Hyperdispatch
  – In V10/V11 that means system engines may wait 3.2ms
  – zIIPs usage should in the 30-50% CPU busy range on average (peaks higher)
• ZAAPZIIP = YES|NO (IEASYSxx option)
  – Allows zAAP eligible workload to run on a zIIP
• zAAP has other settings not applicable to zIIP
  – IFACrossover – disallow zAAP work on general CP

** Be careful about attempting to FORCE zIIP offload
Offload APARS

• PM12256 – zIIP offload improvement up to 55-60%, and less overhead

• PM28626 – corrected inconsistent response time from long running queries after the application of PM12256

• OA35146 – z/OS for PM28626, allows pre-emptible SRB to join/leave and enclave
Results of zIIP Maint.

• Pre- PM12256

<table>
<thead>
<tr>
<th>CLASS 2 TIME DISTRIBUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
</tr>
<tr>
<td>SECP</td>
</tr>
<tr>
<td>NOTACC</td>
</tr>
<tr>
<td>SUSP</td>
</tr>
</tbody>
</table>

• After – PM12256

<table>
<thead>
<tr>
<th>CLASS 2 TIME DISTRIBUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
</tr>
<tr>
<td>SECP</td>
</tr>
<tr>
<td>NOTACC</td>
</tr>
</tbody>
</table>

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

- **II12836** - Info APAR for parallelism V6- DB2 9
- **II14219** - zIIP Exploitation
- **OA38829** - z/OS V1.12 and z/OS V1.13 remove the restriction that prevents zAAP-eligible workloads from running on zIIP processors when a zAAP is installed on the server
- **OA37201** - faster switch to SRB mode
- **PM62824** – DFSORT offload of memory object workfile sorting to zIIP
- Techline Sizing with zCP3000 tool: contact your local IBMer
- RMF Spreadsheet Reporting Tool
- Getting Started Resources
  - [http://www-03.ibm.com/systems/z/hardware/features/ziip/resources.html](http://www-03.ibm.com/systems/z/hardware/features/ziip/resources.html)
- Link to article on ZIIPMAXQL
- Reduce your software costs with IBM System z Integrated Information Processor (zIIP)
Questions???

• 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