POWER8 for DB2 and SAP

Walter Orb
IBM SAP Competence Center, Walldorf

© 2014 IBM Corporation
Agenda

- OpenPOWER Foundation
- POWER8
- POWER8 for SAP
- POWER8 for DB2
Important Disclaimer

IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.

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 is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including 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 results similar to those stated here.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer. Nothing contained in these materials is intended to, nor shall have the effect of, stating or implying that any activities undertaken by you will result in any specific sales, revenue growth or other results.

Some of the information in this document is proprietary to SAP and copyrighted by SAP. No part of this information may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG.
Trademarks

© Copyright IBM Corporation 2014. All rights reserved.

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

IBM, the IBM logo, ibm.com, AIX and DB2 are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
Windows is a trademark of Microsoft Corporation in the United States, other countries, or both.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Other company, product, or service names may be trademarks or service marks of others.

SAP, SAP NetWeaver, SAP Business Information Warehouse, SAP BW, SAP NetWeaver BW, SAP ERP and other SAP products and services mentioned herein are trademarks or registered trademarks of SAP AG in Germany and in several other countries.
Google evaluates Power as alternative for their data centers.
Tyan

Tyan POWER8 motherboard
NVIDIA

NVIDIA Announces Tesla K40 GPU Accelerator and IBM Partnership In Supercomputing

IBM Partners with NVIDIA to Build Next-Generation Supercomputers

GPU-Accelerated POWER-Based Systems Available in 2014
The IBM-backed OpenPOWER Foundation has taken a major step forward this week: Two Chinese organizations, Suzhou PowerCore Technology Company and the Research Institute of Jiangsu Industrial Technology (RIJIT), will join the OpenPOWER Foundation. More importantly, Suzhou PowerCore plans to license IBM’s POWER architecture, intellectual property related to POWER8, and chip design tools to develop and market processors for specialized servers in China. The Research Institute of Jiangsu...
FPGA (field-programmable gate array)

- Xilinx Demonstrates FPGA-Based Acceleration Technology for Next-Generation Data Centers at IBM Impact 2014
- As a member of the IBM OpenPOWER Foundation, Xilinx delivers industry's first key value store accelerator demo based on the IBM CAPI protocol
- Push towards Linux on Power little endian
  - SUSE® Linux Enterprise Server 12 on IBM POWER8
  - Ubuntu for POWER8
  - Little endian support has opened the possibility to optimize the ABI.
- Same virtualization as known from AIX now available for Linux on Power
  - PowerKVM
  - PowerVC
  - PowerVM
- Advanced Toolchain (gcc and others) POWER8 ready
- DB2 LUW currently available for Linux on Power big endian (SuSE + Redhat)
- SAP supports currently Linux on Power big endian
POWER8

© 2014 IBM Corporation
Caches
- 512 KB SRAM L2 / core
- 96 MB eDRAM shared L3
- Up to 128 MB eDRAM L4 (off-chip)

Memory
- Up to 230 GB/s sustained bandwidth

Bus Interfaces
- Durable open memory attach interface
- Integrated PCIe Gen3
- SMP Interconnect
- CAPI (Coherent Accelerator Processor Interface)

Cores
- 12 cores (SMT8)
- 8 dispatch, 10 issue, 16 exec pipe
- 2X internal data flows/queues
- Enhanced prefetching
- 64K data cache, 32K instruction cache

Accelerators
- Crypto & memory expansion
- Transactional Memory
- VMM assist
- Data Move / VM Mobility

Energy Management
- On-chip Power Management Micro-controller
- Integrated Per-core VRM
- Critical Path Monitors

Technology
- 22nm SOI, eDRAM, 15 ML 650mm2
**POWER8 Core**

**Execution Improvement vs. POWER7**
- SMT4 → SMT8
- 8 dispatch
- 10 issue
- 16 execution pipes:
  - 2 FXU, 2 LSU, 2 LU, 4 FPU, 2 VMX, 1 Crypto, 1 DFU, 1 CR, 1 BR
- Larger Issue queues (4 x 16-entry)
- Larger global completion, Load/Store reorder
- Improved branch prediction
- Improved unaligned storage access

**Core Performance vs. POWER7**
- ~1.6x Single Thread
- ~2x Max SMT

**Larger Caching Structures vs. POWER7**
- 2x L1 data cache (64 KB)
- 2x outstanding data cache misses
- 4x translation Cache

**Wider Load/Store**
- 32B → 64B L2 to L1 data bus
- 2x data cache to execution dataflow

**Enhanced Prefetch**
- Instruction speculation awareness
- Data prefetch depth awareness
- Adaptive bandwidth awareness
- Topology awareness
**POWER8 CAPI**  
Coherent Accelerator Processor Interface

**Virtual Addressing**
- Accelerator can work with same memory addresses that the processors use
- Pointers de-referenced same as the host application
- Removes OS & device driver overhead

**Hardware Managed Cache Coherence**
- Enables the accelerator to participate in “Locks” as a normal thread
- Lowers Latency over IO communication model

**Customizable Hardware Application Accelerator**
- Specific system SW, middleware, or user application
- Written to durable interface provided by PSL

**Processor Service Layer (PSL)**
- Present robust, durable interfaces to applications
- Offload complexity / content from CAPP

**POWER8 Coherence Bus**

**PCIe Gen 3**
Transport for encapsulated messages
Coherent Accelerator Processor Interface (CAPI) Overview

Typical I/O Model Flow

Flow with a Coherent Model

Advantages of Coherent Attachment Over I/O Attachment

- Virtual Addressing & Data Caching
  - Shared Memory
  - Lower latency for highly referenced data
- Easier, Natural Programming Model
  - Traditional thread level programming
  - Long latency of I/O typically requires restructuring of application
- Enables Apps Not Possible on I/O
  - Pointer chasing, etc...

© 2014 IBM Corporation

IBM Confidential Until Announce on April 23, 2014 at 1 PM Eastern
Introducing...A NEW GENERATION OF IBM Power Systems

- **Designed for Big Data**
- **Superior Cloud Economics**
- **Open Innovation Platform**

### Power S822L
- 1 or 2 sockets
- 10 or 12 cores/socket
- Up to 1 TB of Memory

### Power S812L
- 1 or 2 sockets
- 6, 8, 10 or 12 cores/socket
- Up to 1 TB of Memory

### Power S824 or Power S814
- 1 or 2 sockets
- 6, 8, 10 or 12 cores/socket
- Up to 1 TB of Memory

### Power S822
- 1 or 2 sockets
- 6, 8, 10 or 12 cores/socket
- Up to 1 TB of Memory

**Linux**

**AIX**

**IBM Cloud Foundry**

**Ubuntu**

**Red Hat**

**SUSE**

© 2014 IBM Corporation
Power S822

*Scale-out application server for secure infrastructure built on open technology*

- 1 or 2 socket, 2 rack units with 6 or 10-core processor card,
- 16-1TB memory
- PowerVM virtualization
- AIX, RHEL, SUSE OS
- Consolidation of UNIX and x86 Linux workloads
- Increase Capacity
  - 12 SFF Bays -- Split Backplane: 6 + 6
  - or 8 SFF Bays & 6 1.8” SSD Bays with Easy Tier with 7GB write cache
- Nine concurrent maintenance PCIe Gen3 slots
  - Higher throughput
  - SR-IOV capable I/O (SOD)
  - Support one CAPI per socket
- Up to 32 disk expansion units
- Greater RAS
  - Standard, split, and high perf RAID backplanes
  - RAID 0, 5, 6, 10 in the base
  - Hot Plug PCIe Gen3 slots
  - Standard redundant power
POWER8 for SAP
Selected Power System Models – SAPS Scalability
Top 20 SAPS/core Servers - Clearly Dominated by IBM Power Systems

DB2 Already Exploits Unique Features of POWER Core

- POWER6 on core Decimal Floating Point Unit
- DB2 is the only DBMS vendor to natively support DECFLOAT data type
  - Performance advantage for retail and finance
  - 40% performance gain in SAP BW
  - Have seen up to 6x faster performance
- DECFLOAT is internally used by DB2 Oracle compatibility mode for NUMBER data type.
  - DB2 Oracle compatibility is used for example for SAP Identity Management product (SAP IDM)
- POWER7 features used by BLU
  - Decimal Floating Point Unit (DFU) utilization for DECIMAL data type conversions
  - VSX vector unit
  - POWER7 instruction set fully exploited by DB2 10.5 BLU
AIX uses the POWER7+ accelerators for
- Encrypted File System (EFS)
- Internet Protocol Security (IPSec)
- /dev/random (HW based random number generator)
- PKCS11 (Public Key Cryptography Standards)
- Active Memory Expansion (requires AIX 6.1 TL4 SP2)
  - Speeds up the compression of data in memory
  - Enables higher compression rates:
    25% more compression in comparison to POWER7

POWER7+ Features used by AIX
Backup and Log File Compression on POWER7+

On-going investigation to use NX842 compression for DB2 backups and log files

Advantages:

- Compression / decompression on accelerator without high CPU utilization
- NX842 compression is extremely fast

Disadvantage:

- Slightly worse compression rate
DB2 Backup/Restore size/time

Comparison NX842 Compression vs. DB2 Standard Compression

- Blue: Runtime Backup to /dev/null in relation to uncompressed
- Red: Size in relation to uncompressed
- Green: Runtime restore in relation to uncompressed

Uncompressed | Standard compression | NX842 HW compression
CPU Usage during backup

Lab Tests

CPU Total

Uncompressed  DB2 Standard Compression  NX842 Compression
Comparison Standard Compression vs. ZLib FPGA Accelerated Compression

- **Runtime Backup to /dev/null in relation to uncompressed**
- **Size in relation to uncompressed**
- **Runtime restore in relation to uncompressed**
POWER8 features for DB2 without software change

- Faster DECFLOAT operations
  - Faster conversion from internal BLU DECIMAL format to DB2 BCD format and vice versa
- Faster integer divide
- Doubled vector unit for faster column scans in BLU
- SMT-8
  - more throughput per core
- More throughput through
  - Faster memory access
  - Bigger caches
Planned POWER8 exploitation in DB2 – row based DECIMAL SUM

Extract from POWER ISA 2.07

**Decimal Add Modulo VX-form**

```
bodadd: VRT,VRA,VRB,PS

<table>
<thead>
<tr>
<th></th>
<th>4</th>
<th></th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>11</td>
<td>18</td>
<td>21</td>
<td>3</td>
</tr>
</tbody>
</table>
```

if MSR.WEC=0 then Vector_Unavailable()

VR[VRT] ← Signed_BCD_Add(VR(VRA),VR(VRB),PS)

CR.bit[56] ← inv_flag ? 0 : lt_flag
CR.bit[57] ← inv_flag ? 0 : gt_flag
CR.bit[58] ← inv_flag ? 0 : eq_flag
CR.bit[59] ← ox_flag | inv_flag

**Lab Tests**

![Faster Decimal Math Processing](image)

11% Improved!
Planned Power8 exploitation in DB2 – BLU column scan

Lab Tests

Detailed analysis of BLU scan performance. Shows the value of new instructions added to POWER8 vector unit (VSX).
Planned POWER8 exploitation in DB2 – Page checksum Lab Tests

**Faster Integrity Checks**

- **38% Improved!**

Comparison between Power 8 Vectorized Checksum and Original Checksum.
Effect of SIMD optimizations

<table>
<thead>
<tr>
<th>Evaluator</th>
<th>SIMD=0 All Off</th>
<th>Evaluator Time (ms)</th>
<th>SIMD=1 All On</th>
<th>Evaluator Time (ms)</th>
<th>Ratio to All Off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluator 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluator 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only available if you use POWER7 or higher! ➔ Use POWER7 or higher for BLU. On Intel, SSE is used to accelerate BLU.
DB2 exploits the POWER architecture

POWER

- POWER on core Decimal Floating Point Unit
- DB2 is the only DBMS vendor to natively support DECIMAL data type
  - Performance advantage for retail and finance
  - 40% performance gain in SAP BW
  - Have seen up to 6x faster performance
- POWER7 chip has on-chip L2 / L3 cache with eDRAM L3 Cache.
  - DB2 is cache aware
  - Optimizes power of the core
- POWER7 features used by BLU
  - Decimal Floating Point Unit (DFU) utilization for DECIMAL data type conversions
  - VSX vector unit
  - POWER7 instruction set fully exploited by DB2 10.5 BLU

New in POWER8

- Faster integer divide
- Faster DECFLOAT operations
- Faster conversion from internal BLU DECIMAL format to DB2 BCD format and vice versa
- Doubled vector unit for faster column scans in BLU
- More throughput per core due to SMT8

- DB2 leverages new POWER8 SIMD for range predicates to evaluate many more column values simultaneously compared to Power7 or Intel
- Cognitive compilation
  - In compiling and optimizing DB2 runtime code, IBM uses special cognitive algorithms that watch DB2 processing BLU Acceleration workloads
  - This learning is then used to reorder instructions within the product for even faster runtime performance
Power articles

- Power on HANA:
  http://saponpower.wordpress.com/2014/06/06/there-is-hop-for-hana-hana-on-power-te-program-t

- OpenPower for Chinese Market:

- CAPI article:
  http://dancingdinosaur.wordpress.com/tag/coherent-accelerator-processor-interface-capi-power8/

- Tyan Power8 Motherboard article:
  http://www.enterprisetech.com/2014/04/28/inside-google-tyan-power8-server-boards/

- NVIDIA for Super Computing article:

- Xilinx and OpenPOWER:

- Ubuntu for Power8:
  https://insights.ubuntu.com/2014/06/13/ubuntu-and-power8-the-best-of-both-worlds/

- SLES 12 on Power8:
More Information

- DB2 on Power Resources on IBM.COM – (LINK)
- IBM POWER Roadmap – (LINK)
- SAP Benchmark publications – (LINK)
- Power System Performance: http://www-03.ibm.com/systems/power/hardware/reports/system_perf.html
- OpenPOWER http://openpowerfoundation.org/
- Power8 Architecture: https://www.power.org/documentation/power-isa-version-2-07/