IBM Information On Demand 2008

IBM

Taming a Terabyte of XML Data with DB2® on Intel® Xeon® Processor 7400 Series

Agustin Gonzalez, Intel

Matthias Nicola, IBM Silicon Valley Lab

Session 2623

INNOVATE.OPTIMIZE.PERFORM.

IBM INFORMATION ON DEMAND 2008
October 26 – 31, 2008
Mandalay Bay
Las Vegas, Nevada

Agenda

- Motivation & Use Cases
- DB2® pureXML™
- TPoX Benchmark
- Intel® Xeon® Processor 7400 Series
- Performance Results
  - Mixed Workload: 70% Query, 30% IUD
  - Insert Performance
- Summary

Motivation

- XML: Standard format for SOA and message-based transaction processing

- Common Requirements
  - Persist, index, query, validate and update XML messages with full ACID etc. as in traditional databases
  - Retain and manage Millions of XML messages
  - Multi-TeraByte XML databases

- XML often considered big & slow
  - Based on past experience with insufficient technology
  - No longer true with state-of-the-art database and processor technology

Example: Financial Application Logging

- Customers: Banking and Insurance
- Requirement: Record "everything" that's happening, e.g. in online banking, investment, insurance applications
  - Trace every step a customer takes, by date/sessionID/userID etc.
  - Every step in a web dialog creates a small XML document
  - Functional: application level monitors, used by call-centers, audits, etc.
  - Technical: system and infrastructure trouble shooting, used by IT
- Data Characteristics
  - "Log records" in XML format, common header, *highly variable body which cannot be represented in relational format*
  - High insert rates: 10M – 50M documents per day
  - Accumulating 300M to 1.5B documents per month
    → **Terabytes of XML Data**

Other Customer Requirements

- Large North American Bank
  - SOA services persist 2 – 4 TB of XML messages/day
  - Currently stored as flat files, not queryable
  - Performance and scalability problems
  - Need to store, index and query XML data efficiently

- Non-US Government Agency
  - Requirement to store and index 100M to 1B small XML messages per day, keep data for 7 days
  - TB's of XML data, high insert performance required

Why/When XML instead of Relational ?

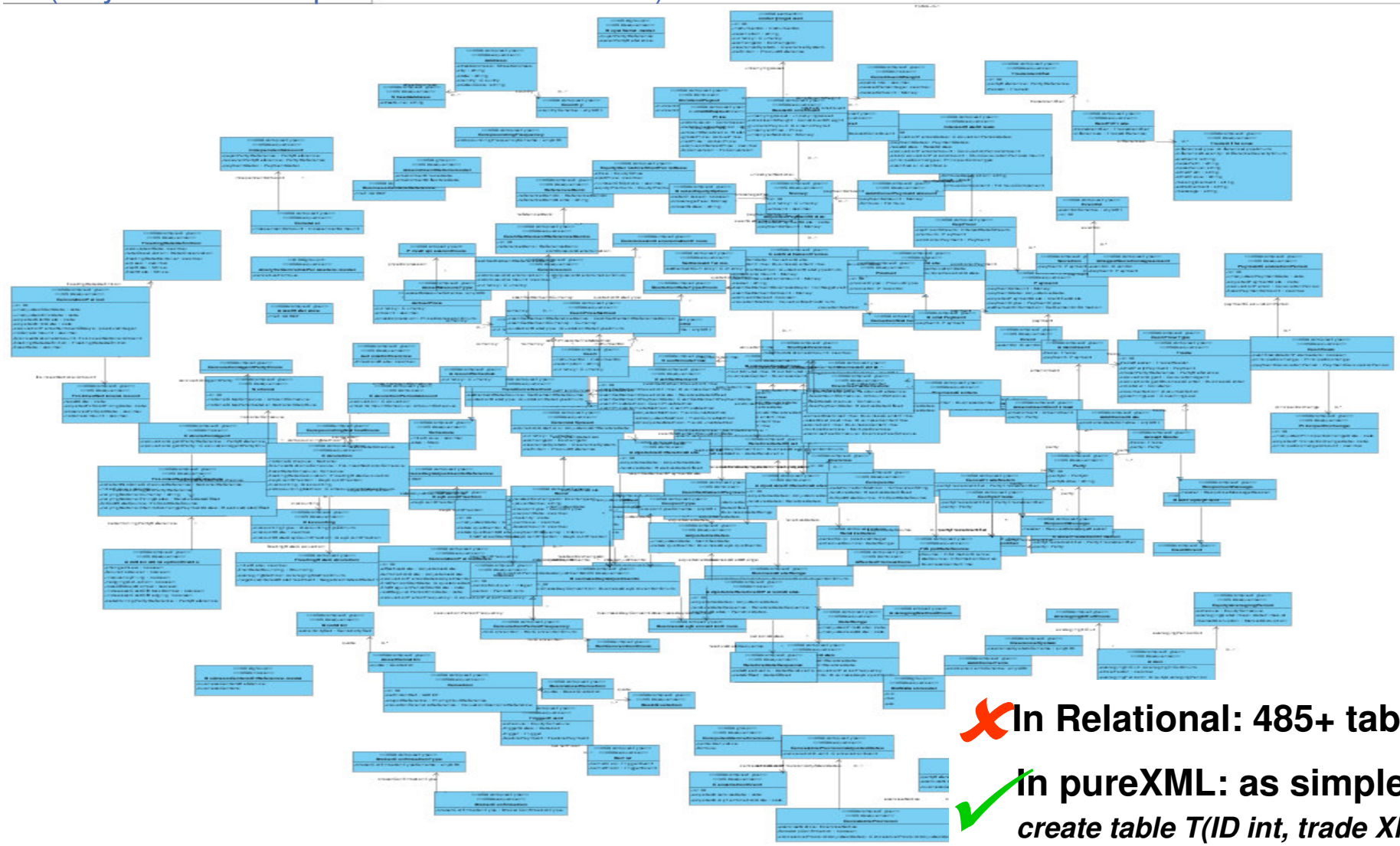
- When data format (schema) changes over time
- When schema is complex and highly variable, such as:

	XML Schema Files	Type Definitions	Elements + Attributes
HL7 CDA 3 Health Level 7, Clinical Document Architecture	6	1953	945 + 477
STAR Standards for Technology in Automotive Retail (OAGIS)	192	5846	77319 + 625
FpML 4.2: Financial products Markup Language	23	686	1867 + 196
FIXML 4.4: Financial Information eXchange Protocol (*used in the TPoX benchmark !*)	41	1310	619 + 2593

- Storage as LOB? → Data cannot be indexed or queried !

Generate Relational Schema for FpML

(only 10% of the FpML schema used here)



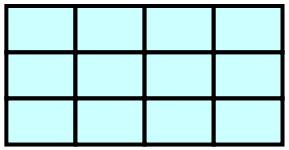
✗ In Relational: 485+ tables

✓ In pureXML: as simple as

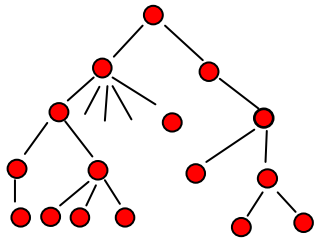
create table T(ID int, trade XML);

1TB of XML vs. 1TB of Relational Data

- Relational



  - Data is highly structured, very regular in nature, rows are flat
  - Fixed Meta Data, fixed number of columns per table
  - Fixed data type for all values in a column
  - A cell in a table holds a single atomic data value
  - Data format defined by DDL, known at query/insert/update compile time
- XML



  - Semi-structured, data highly variable in nature
  - No fixed format, not fixed data types, meta data is variable
  - Data is hierarchical, can be arbitrarily nested
  - XML Document = complex object containing many values
  - Data format not predefined, not known until query/insert/update run time

→ Processing XML is inherently more complex

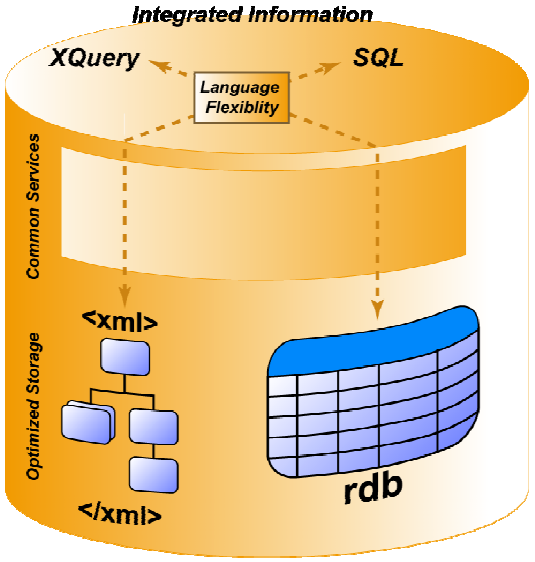
(inserts, indexing, queries, updates, statistics, optimizations,…)

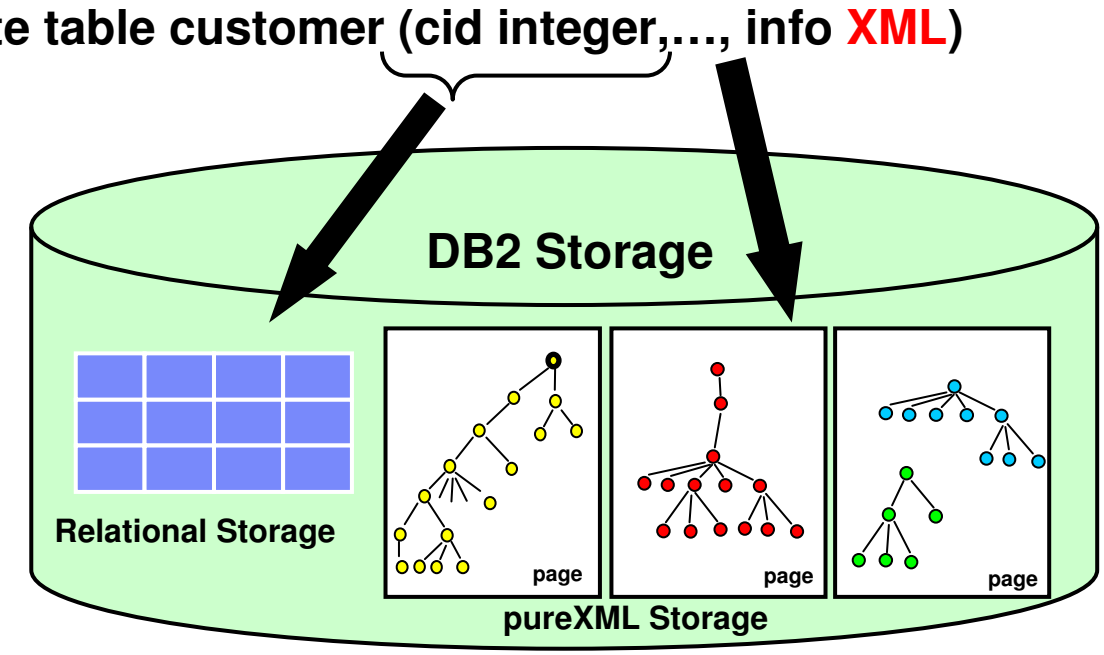
Agenda

- Motivation & Use Cases
- DB2® pureXML™
- TPoX Benchmark
- Intel® Xeon® Processor 7400 Series
- Performance Results
  - Mixed Workload: 70% Query, 30% IUD
  - Insert Performance
- Summary

Overview of DB2 pureXML (1 of 2)

- XML stored in a parsed hierarchical format
- No parsing for XML queries or updates → Performance !
- XML Schema validation is optional, per document
- XML indexes for specific elements/attributes
- XQuery and SQL/XML Integration





Overview of DB2 pureXML (2 of 2)

① create table customer (cid integer, info XML)

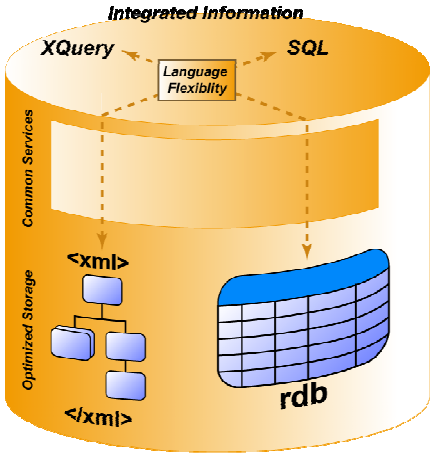
② insert into customer (cid, info) values (?,?)

③ select cid, info from customer

④ select **xmlquery**('$INFO/customer/name')
from customer
where **xmlexists**('$INFO/customer/addr[zip = 95123]')

⑤ create index idx1 on customer(info) generate keys using xmlpattern '/customer/addr/zip' as sql varchar(12)

⑥ Plus: updates, XML Schema support, utilities, etc.



Agenda

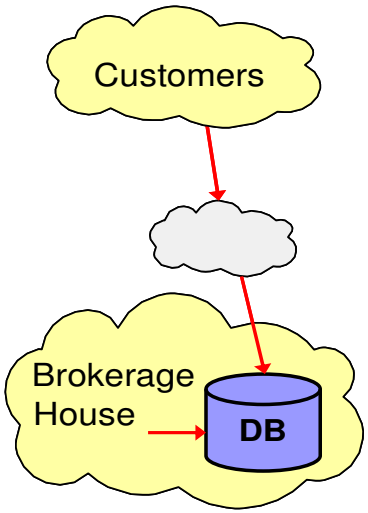
- Motivation & Use Cases
- DB2® pureXML™
- TPoX Benchmark
- Intel® Xeon® Processor 7400 Series
- Performance Results
  - Mixed Workload: 70% Query, 30% IUD
  - Insert Performance
- Summary

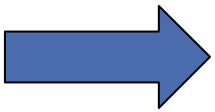
TPoX Benchmark

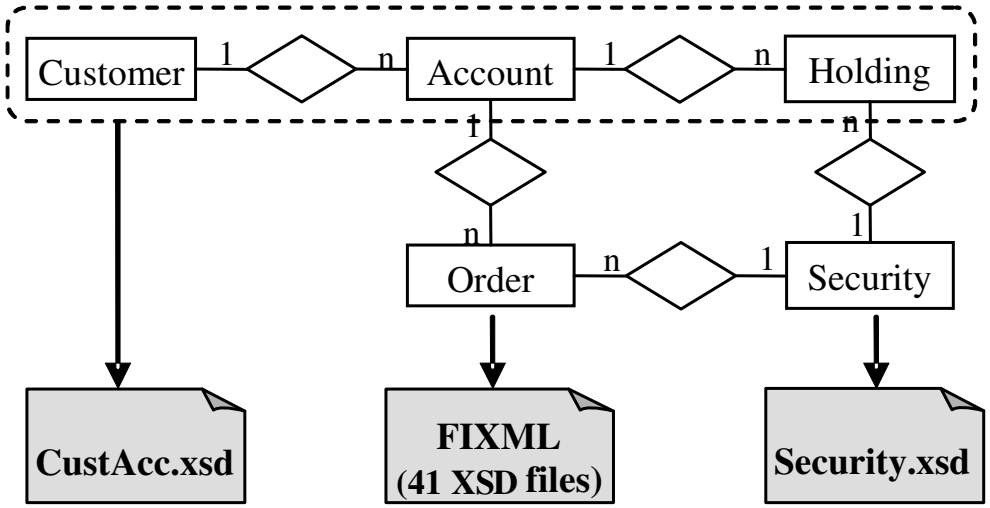
TPoX = **T**ransaction **P**rocessing **o**ver **X**ML Data

Open Source Benchmark: http://tpox.sourceforge.net/

Financial transaction processing scenario: “online brokerage”







4 – 20 kb **1 – 2 kb** **2 – 9 kb**

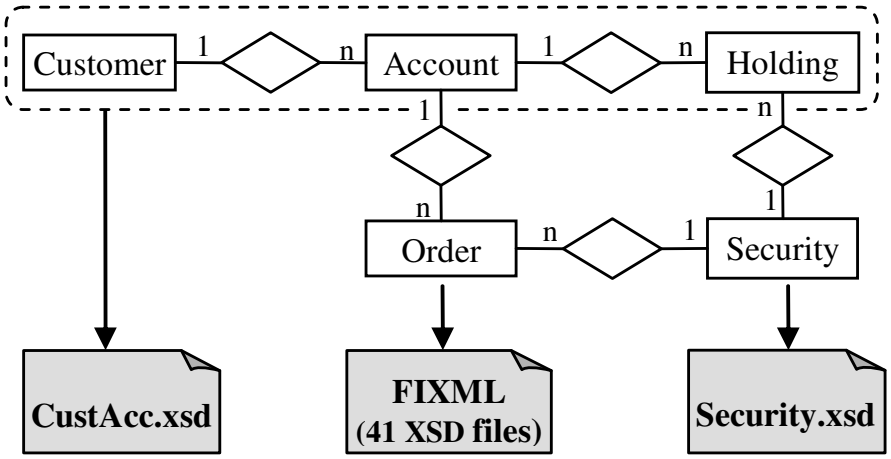
FIXML: Standardized Financial XML Schema for Securities Trading !

TPoX Data & Schema

FIXML: financial industry XML Schema

CustAcc: modeled after a real banking system that uses XML

Security: information similar to investment web sites



- create table custacc (cadoc XML)
- create table security (sdoc XML)
- create table order (odoc XML)

- **Scale Factor "M", 1 TB raw data**
- **300M Order documents, 60M CustAcc documents**
- **20,833 Securities, independent of scale factor**
- **3 Simple Tables + XML Indexes**

TPoX Scalability

Scale	Approx raw size		Security	CustAcc	Orders	Actual Total Raw Data Size
XS	10GB	#Docs:	20,833	600,000	3,000,000	3,620,833
		GB:	0.13	3.62	5.79	9.55
S	100GB	#Docs:	20,833	6,000,000	30,000,000	36,020,833
		GB:	0.13	36.24	57.91	94.28
M	1TB	#Docs:	20,833	60,000,000	300,000,000	360,020,833
		GB:	0.13	362.41	579.07	941.61
L	10TB	#Docs:	20,833	600,000,000	3,000,000,000	3,600,020,833
		GB:	0.13	3624.08	5790.71	9414.92
XL	100TB	#Docs:	20,833	6,000,000,000	30,000,000,000	36,000,020,833
		GB:	0.13	36240.77	57907.10	94148.00
XXL	1PB	#Docs:	20,833	60,000,000,000	300,000,000,000	360,000,020,833
		GB:	0.13	362407.7	579071.0	941480.0

TPoX Workload

Transact.	Purpose	Tables	Weight
Q1	Retrieve an order for a given order id	O	10%
Q2	Retrieve a security for a given ticker symbol w/o root	S	10%
Q3	Get a customer's personal data, construct profile doc.	C	10%
Q4	Search securities based on 4 predicates and return specific elements of interest	S	10%
Q5	Construct an account summary and statement	C	10%
Q6	Retrieve the price of a certain security	S	10%
Q7	Get a customer's most expensive order	C,O	10%
U1	Close an existing customer's account	C	1%
U2	Open a new account for an existing customer	C	1%
U3	Update the price of a security	S	3%
U4	Update the status of an order	O	3%
U5	Execute a "buy" order of a given security for a given account: 1. If shares already exist, increase the quantity; otherwise, add a new holding 2. Replace account balances and values dates 3. Abort if the max. number of holdings is exceeded	C,S	3%
U6	Execute a "sell" order (opposite of U5)	C, S	3%
I1	Customer places a new order (insert order document)	O	7%
I2	Add a new customer (insert CustAcc document)	C	1%
D1	An order is cancelled or archived (delete order doc)	O	7%
D2	Remove a customer (delete CustAcc document)	C	1%

70% Queries

30% I/U/D

TPoX Transaction Characteristics

XML Database Operation	TPoX Transactions
Full document insert/delete	I1, I2 / D1, D2
Full document retrieval	Q2
Element/attribute value update	U3, U4, U5, U6
Subtree insert	U2, U5
Subtree delete	U1, U6
Subtree replace	U5, U6
Element construction	Q3, Q4, Q5, Q6, U2, U5, U6
Predicate evaluation	all 17 transactions
*, // processing	Q4
Join across document types	Q7, U5, U6
Aggregation	Q7
Arithmetic on XML values	Q7, U5, U6
Schema validation required	I2, U2, U4

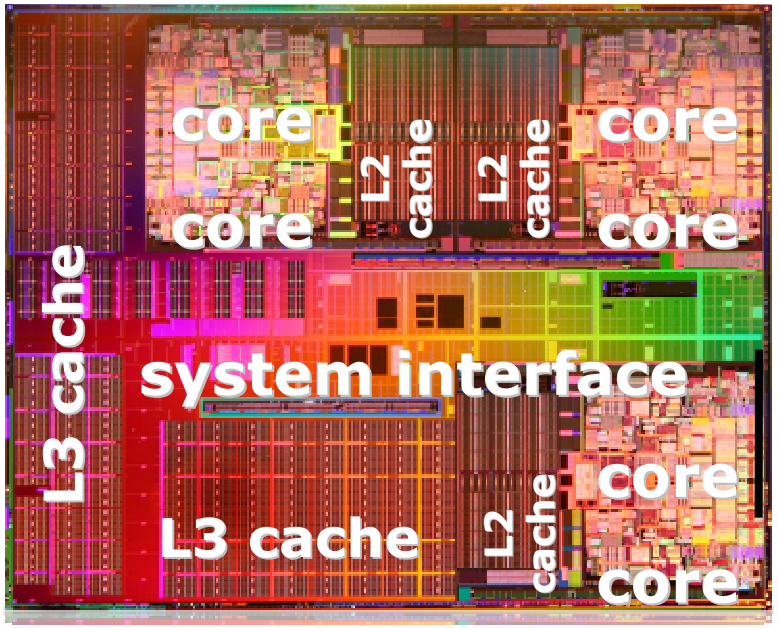
- **Rich workload, broad range of realistic XML operations**

Agenda

- Motivation & Use Cases
- DB2® pureXML™
- TPoX Benchmark
- Intel® Xeon® Processor 7400 Series
- Performance Results
  - Mixed Workload: 70% Query, 30% IUD
  - Insert Performance
- Summary

Intel® Xeon® processor 7400 Series

- New Features
  - 6 cores per socket
  - 45nm Hi-k metal gate process technology
  - Large L3 cache (16MB) shared by all 6 cores
  - Each core pair shares 3M L2 cache
  - 40-bit physical addressing

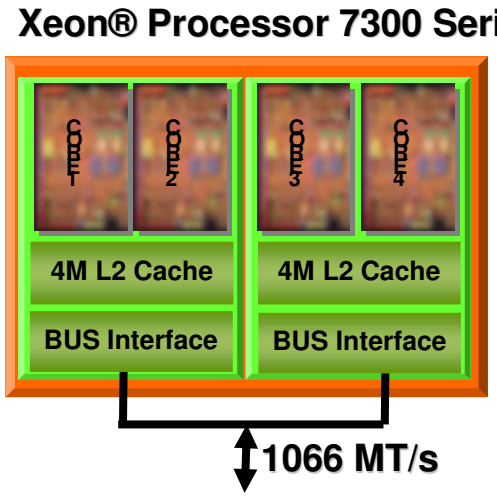


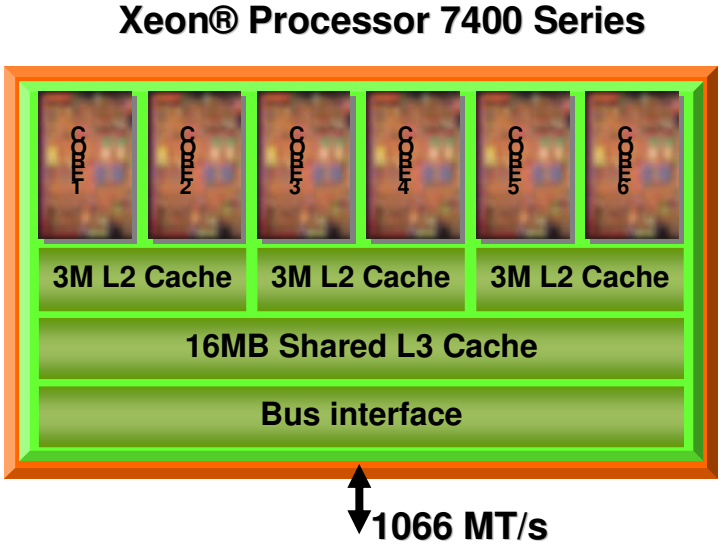
- Continued Features from 2007
  - Intel® Core™ Micro-architecture
  - 1066 MHz Bus Speed
  - Dedicated High-Speed Interconnects
  - Clarksboro chipset

- Mechanical/Thermal
  - Socket mPGA604 (F')
  - Pin compatible with 7200 & 7300 series processors
  - Power 50W to 130W
  - Targeting Rack-Optimized & Ultra-Dense SKUs
  - Supports VT-x (Intel® Flex Migration and Flex Priority)

Intel® Xeon® Processor 7400 Series

Intel® Xeon® Processor 7300 Series vs. Intel® Xeon® Processor 7400 Series





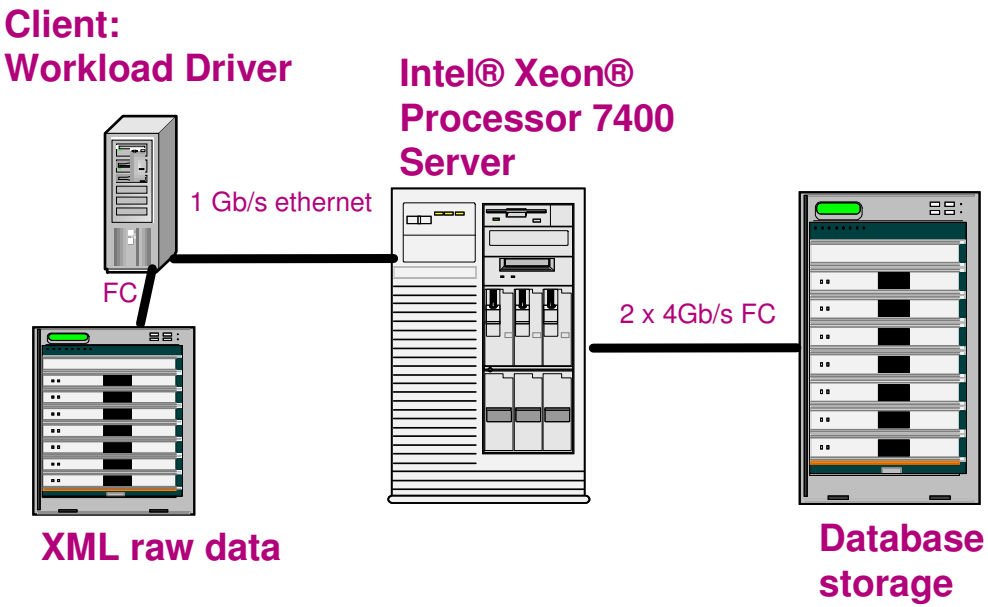
Attribute	Intel® Xeon® Processor 7300 Series	Intel® Xeon® 7400 Processor Series
Cores Per Processor	Up to four cores per processor	Up to six cores per processor
Process technology	65nm	45nm Hi-k
Frequency	Up to 2.93GHz	Up to 2.66GHz
Power	130W/80W/50W	130W/90W/65W/50W
Micro architecture	Intel® Core™ Micro-architecture	Intel® Core™ Micro-architecture
L2 Cache	Up to 4M Per Core Pair – Total 8M L2	3M Per Core Pair– Total 9M L2
L3 Cache	No L3 Cache	Up to 16MB L3 Cache
Platform	Caneland/OEM	Caneland/OEM
Chipset/FSB Speed	Intel® 7300 or OEM Chipset/1066 MHz	Intel® 7300 or OEM Chipset/1066 MHz
Memory	Up to 32 Dimms (Max 256GB)	Up to 32 Dimms (Max 256GB)

Agenda

- Motivation & Use Cases
- DB2® pureXML™
- TPoX Benchmark
- Intel® Xeon® Processor 7400 Series
- Performance Results
  - Mixed Workload: 70% Query, 30% IUD
  - Insert Performance
- Summary

System and Storage Configuration Used

System Under Test:



Software:

- 64-bit SuSE* Linux* Enterprise 10, SP1
- DB2® 9.5 FP2
- TPoX Workload Driver (Java*)
- Performance tools (iostat, sar, etc.)

Platform Name:	Intel® Xeon® Processor 7400 Series
Processors:	4 CPUs 6 cores per CPU 16 MB L3 Clock frequency: 2.67 GHz
Memory:	64 GB/DDR2-667
Storage:	1 Internal Drive + 120 Disk RAID0 Setup contains the DB2 database, 15 Disk RAID0 for logging 30 Disk RAID0 for raw XML data
HBA Interface:	1 PCI-E 4Gb/s Dual Fiber Channel Adapter for Server, 1 PCI-E 4Gb/s Dual FC for Client
BIOS and Chipset:	T136, Intel® 7300 Chipset (Clarksboro)/1067 MHz

DB2 Configuration Used

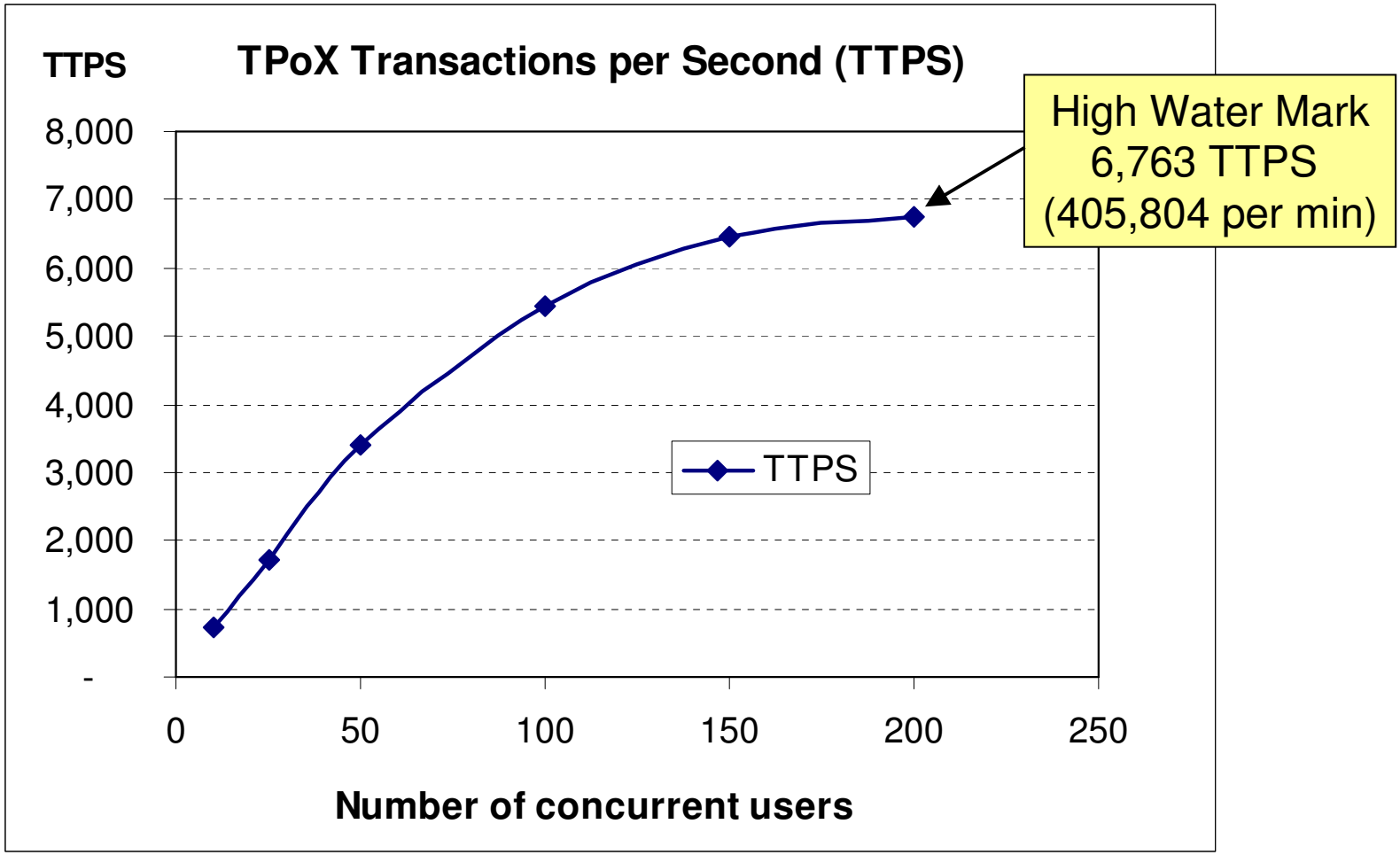
- DB2 9.5 FP2 on Linux* (SLES 10)
- Entire Database: automatic storage, 16kb pages
- 5 Table Spaces + Buffer Pools:
  - order, orderIdx, custacc, custaccIdx, security
  - no file system caching
- XML Inlining and Compression
- Self-tuning memory management for all memory areas: buffer pools, sort heap, lock list, etc.
  - `INSTANCE_MEMORY=automatic`
  - `DATABASE_MEMORY=automatic`
- DB2's "automatic" setting also used for `num_iocleaners` and other performance knobs
- `DB2_USE_ALTERNATE_PAGE_CLEANING=Y`

Benchmark Results: Tables & Indexes

- CustAcc Data (60M XML Documents)*
  - Space Used: 7,959,808 pages → **121.4 GB**
  - Compression Ratio: **64%**
  - Size of Indexes: 674,048 pages → **10.3 GB**
- Order Data (300M XML Documents)*
  - Space Used: 17,643,104 pages → **269.2 GB**
  - Compression Ratio: **57%**
  - Size of Indexes: 2,573,216 pages → **39.3 GB**
- Total Database Size: **440.2 GB***
  - *incl. order, custacc, security tables and all indexes*
  - *remember: raw data volume without indexes is 1 TB !*

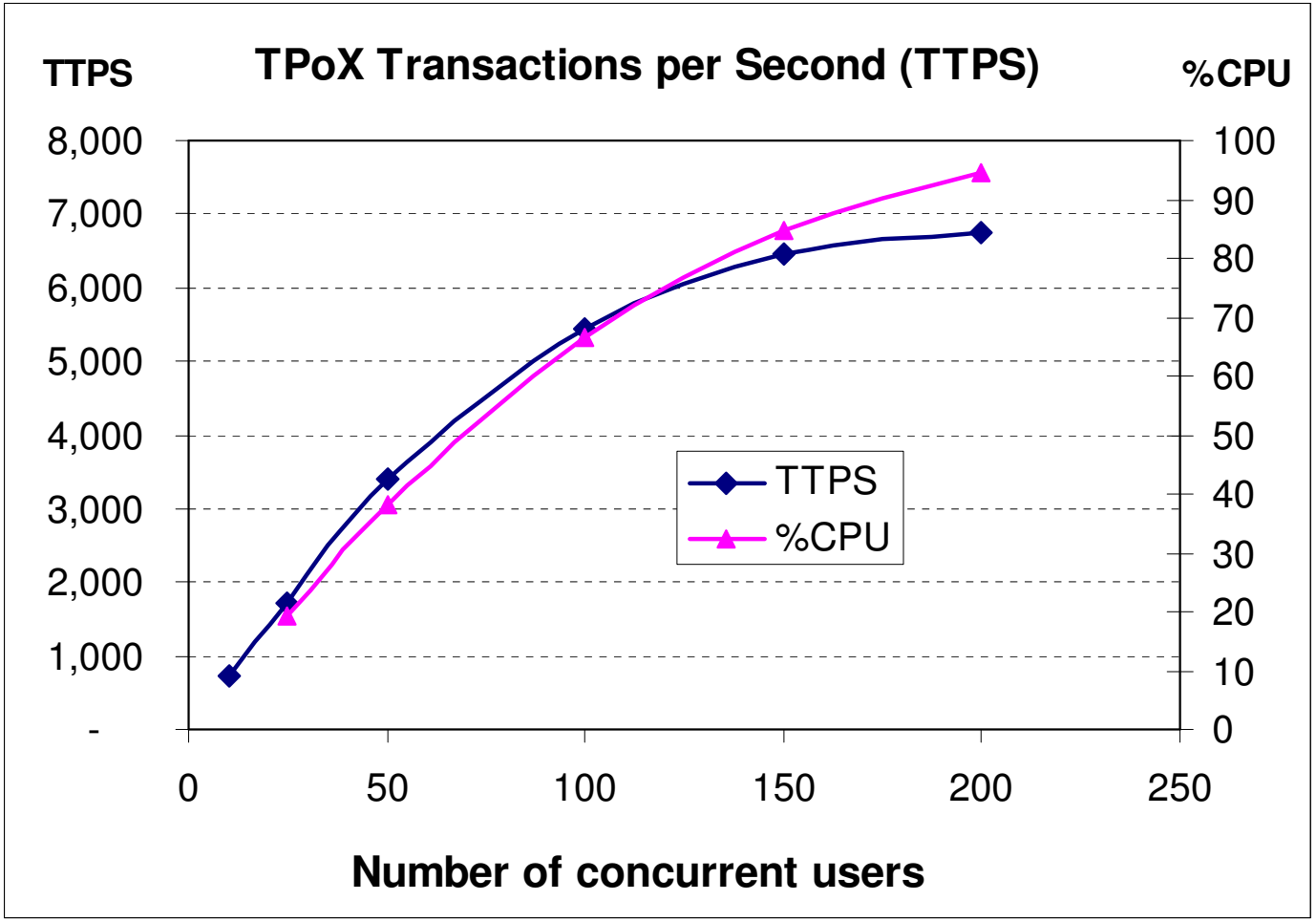
* Source: Intel internally measured results, September 2, 2008

Benchmark Results: Mixed Workload*



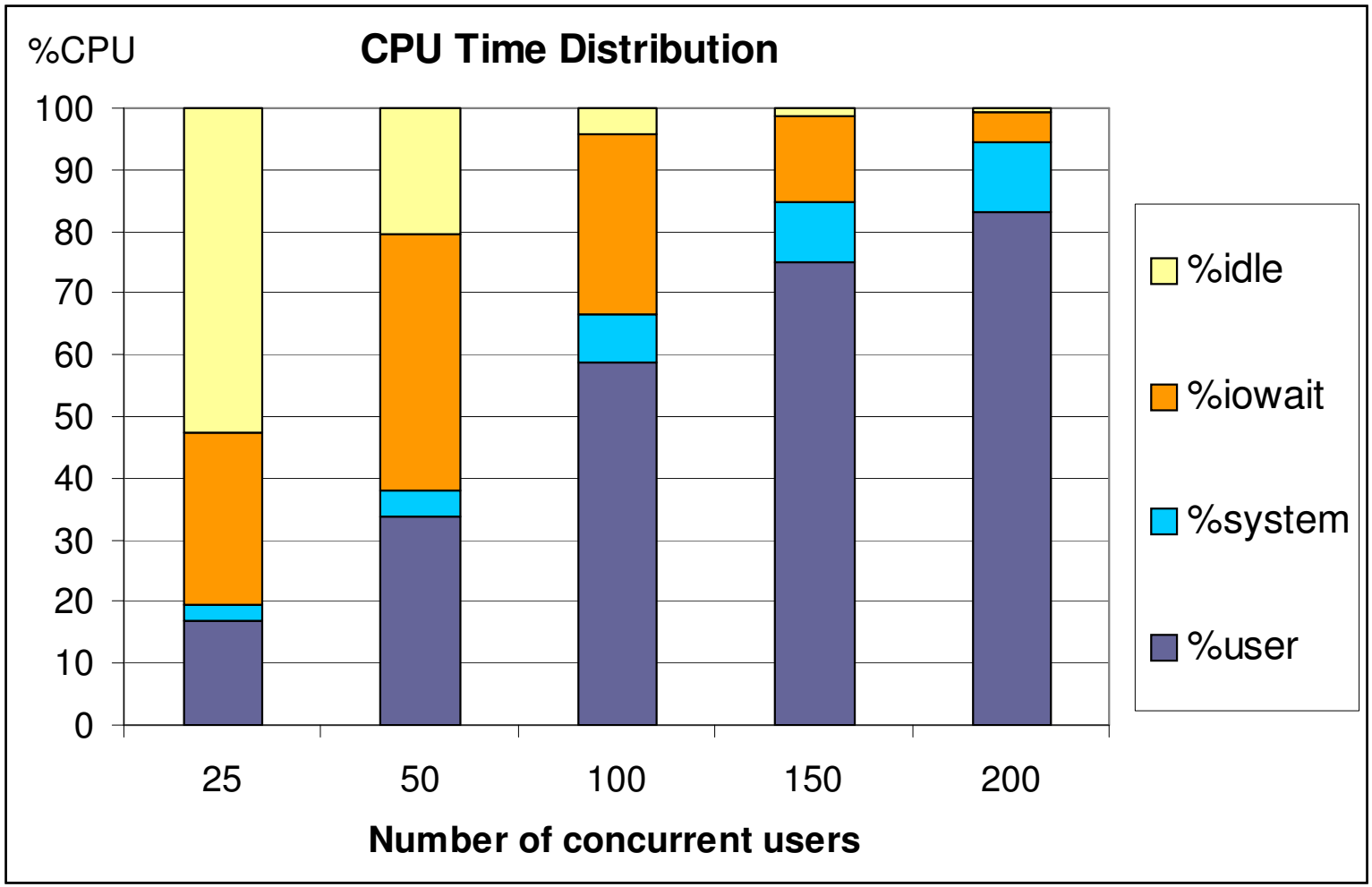
* Source: Intel internally measured results, September 2, 2008

Benchmark Results: Mixed Workload*



* Source: Intel internally measured results, September 2, 2008

CPU Time Distribution*

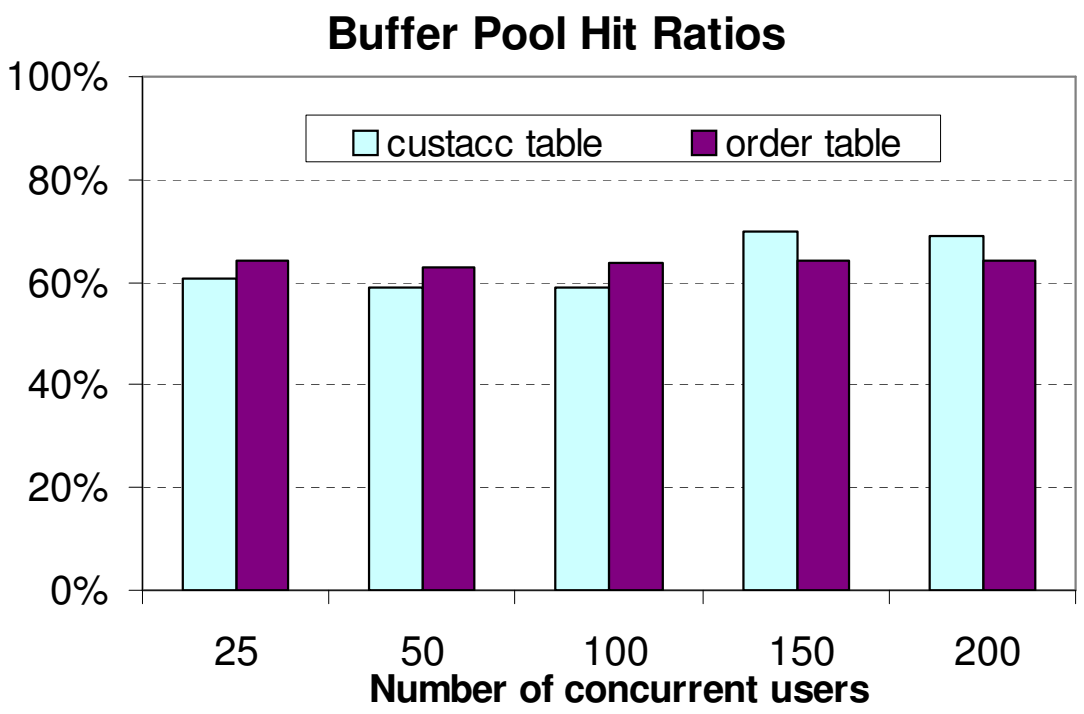


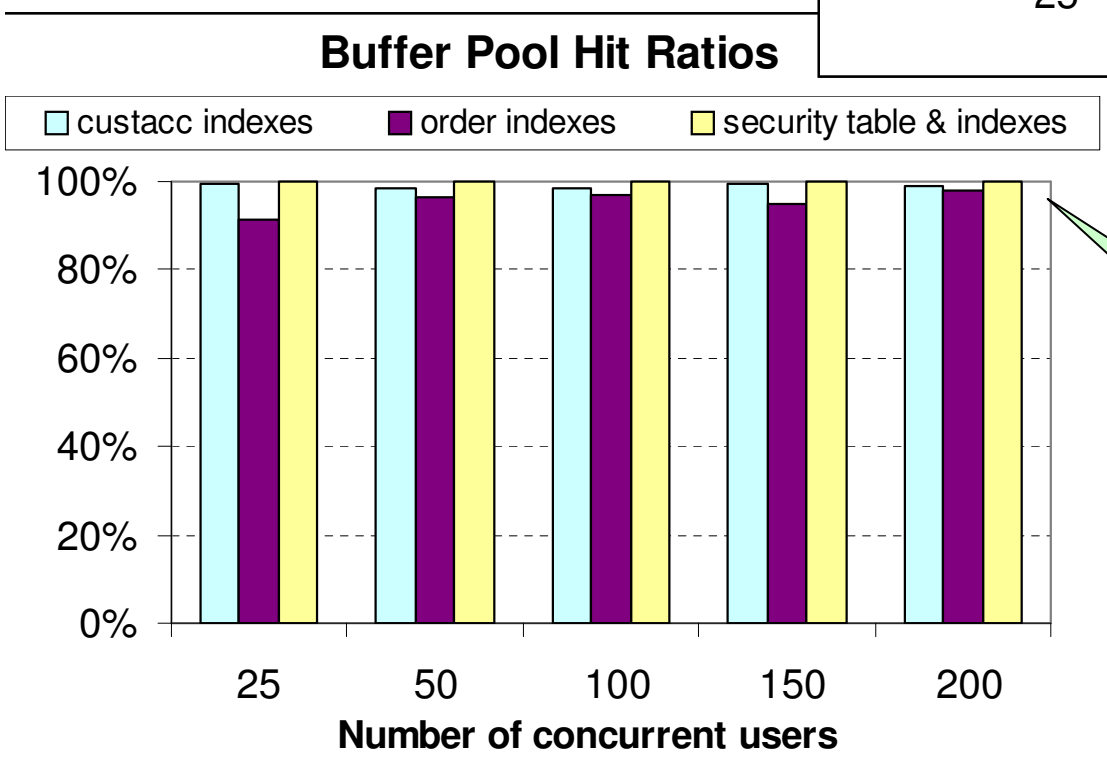
* Source: Intel internally measured results, September 2, 2008

DB2 Buffer Pool Hit Ratios*

- All buffer pool sizes self-tuned by DB2 !

* Source: Intel internally measured results, September 2, 2008

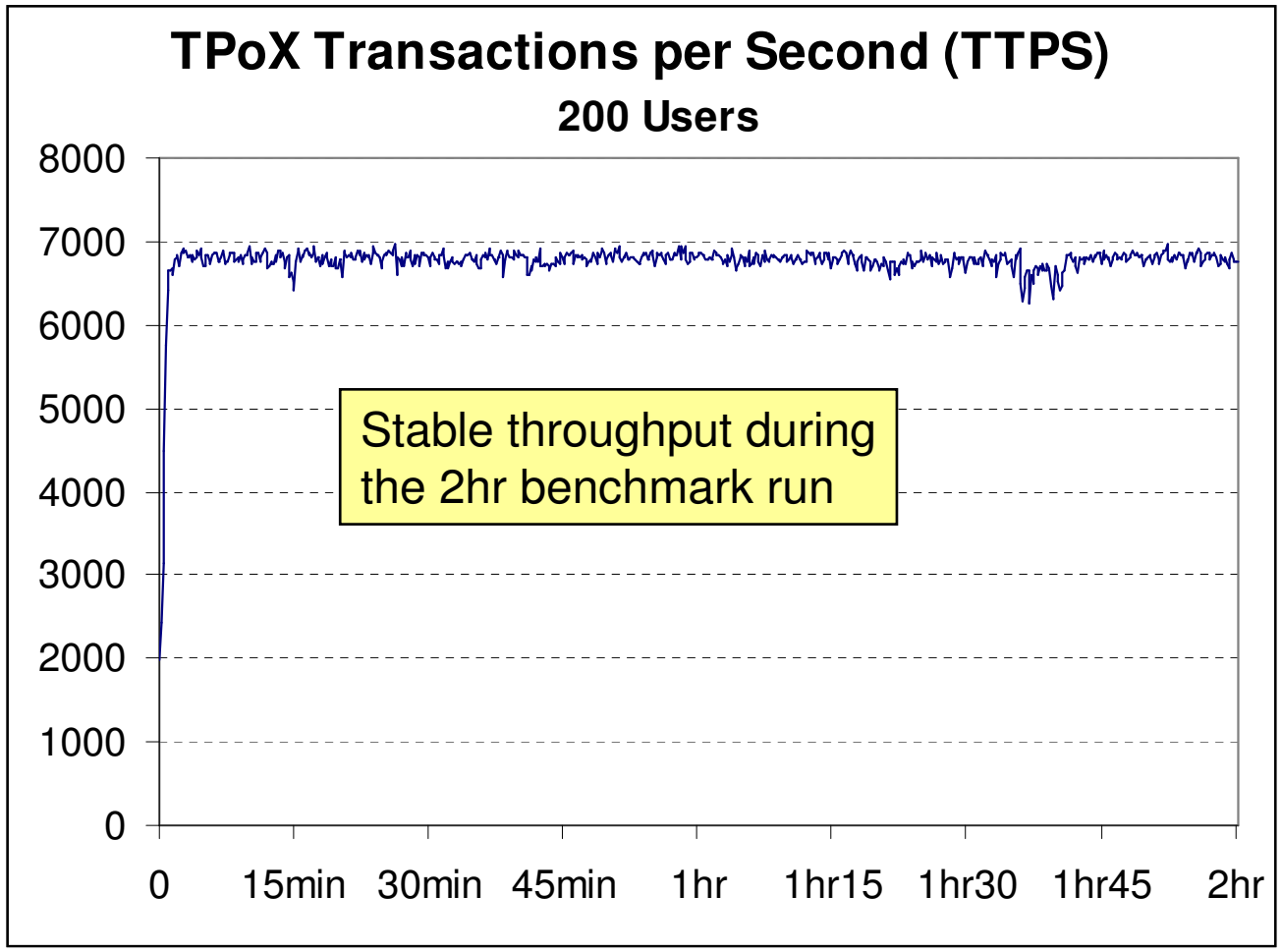




Random I/O into the large tables

>98% hit ratio for all indexes and security table

Mixed Workload at 200 Users*



* Source: Intel internally measured results, September 2, 2008

Mixed Workload at 200 Users *

Queries	34,011,891	70%
Updates	6,806,323	14%
Inserts	3,890,328	8%
Deletes	3,886,554	8%
Total	**48,595,096**	

Output from a 2hr Benchmark Run:

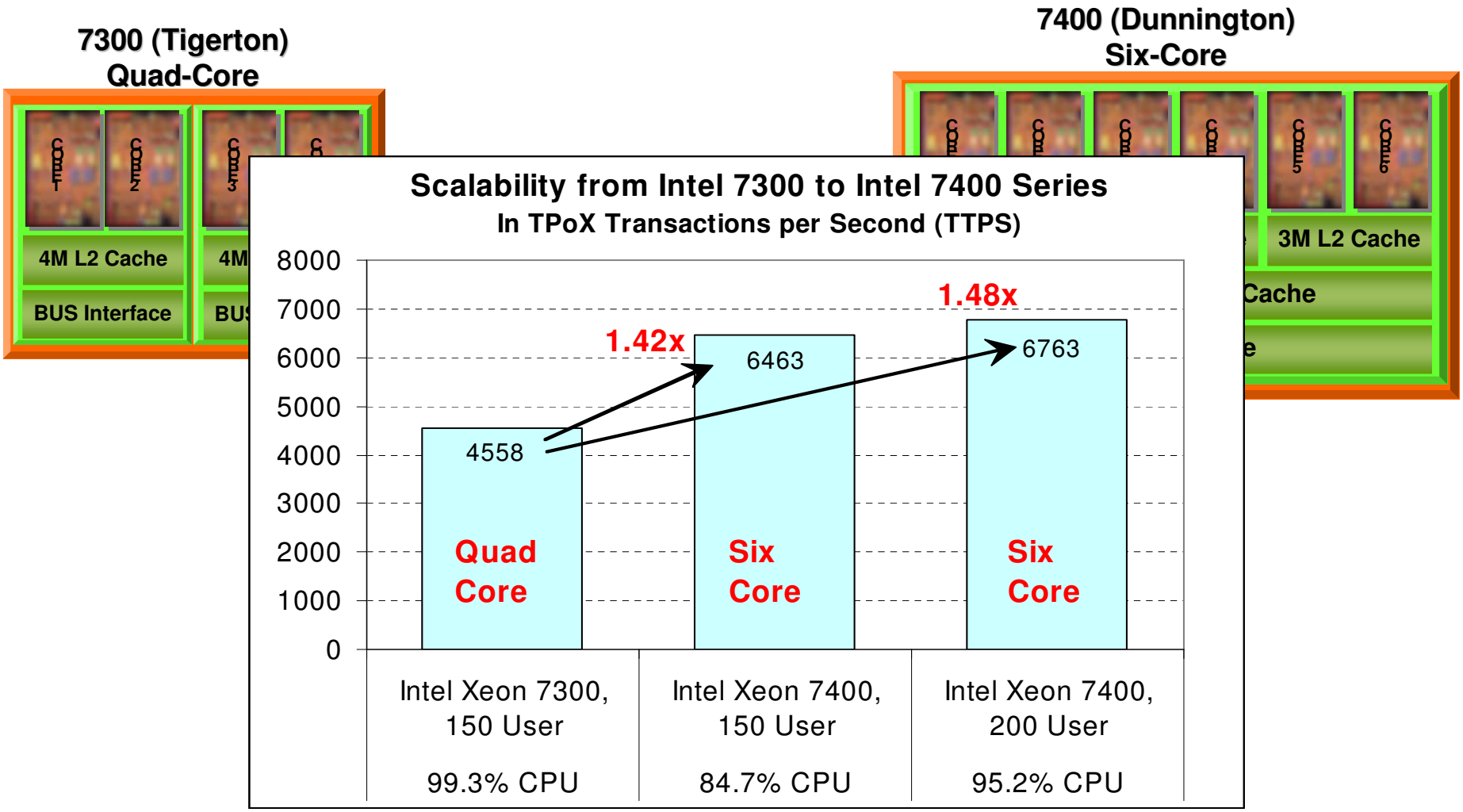
```
*** WORKLOAD STATISTICS ***
```

Tr.#	Name	Type	Count	%-age	Total Time(s)	Min Time(s)	Max Time(s)	Avg Time(s)
1	Get_order	Q	4859631	10.00	84490.77	0.00	0.37	0.02
2	Get_security	Q	4855112	9.99	31999.10	0.00	0.21	0.01
3	Customer_profile	Q	4863296	10.01	79068.06	0.00	0.21	0.02
4	Search_securities	Q	4861991	10.01	286924.50	0.00	0.54	0.06
5	Account_summary	Q	4855457	9.99	86128.72	0.00	0.36	0.02
6	Get_security_price	Q	4859441	10.00	30378.65	0.00	0.15	0.01
7	Customer_max_order	Q	4856963	9.99	253992.34	0.00	0.26	0.05
8	U1CloseAccount	U	485654	1.00	15431.24	0.00	1.68	0.03
9	U2OpenAccount	U	486821	1.00	31283.13	0.00	1.94	0.06
10	U3SecurityPrice	U	1458598	3.00	33801.16	0.00	0.21	0.02
11	U4OrderStatus	U	1460055	3.00	61331.56	0.00	0.58	0.04
12	U5BuySecurity	U	1457954	3.00	55542.71	0.00	1.83	0.04
13	U6SellSecurity	U	1457241	3.00	54253.77	0.00	1.69	0.04
14	delcustacc	D	485762	1.00	14893.65	0.00	0.21	0.03
15	delorder	D	3400792	7.00	105141.35	0.00	0.69	0.03
16	insValidcustacc	I	487083	1.00	14013.69	0.00	1.87	0.03
17	insNoValidorder	I	3403245	7.00	67252.42	0.00	0.93	0.02
			48,595,096					

```
The throughput is 405804 transactions per minute (6763.42 per second).
```

* Source: Intel internally measured results, September 2, 2008

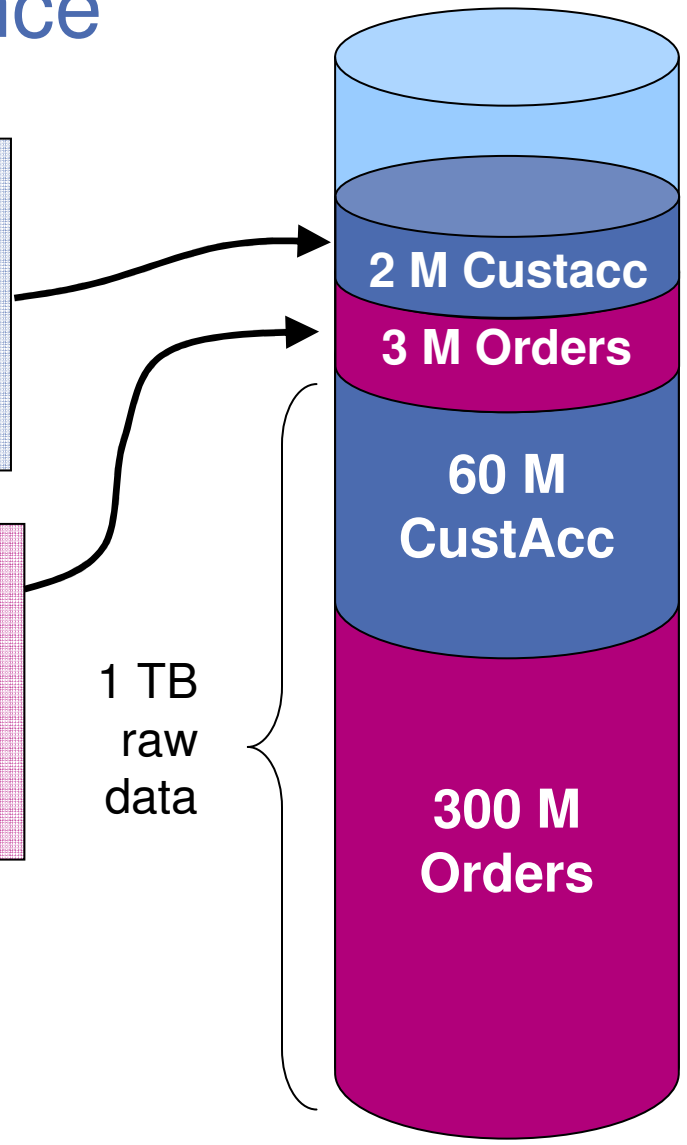
Intel® Xeon® Processor 7300 vs. 7400* Series



* Source: Intel internally measured results, September 2, 2008

Incremental Insert Performance*

- Custacc (4 – 20kb):
  - 4,913 inserts per second
  - 101 GB / hour
- Orders (1 - 2kb):
  - 11,904 inserts per second
  - 69 GB / hour



* Source: Intel internally measured results, September 2, 2008

Agenda

- Motivation & Use Cases
- DB2® pureXML™
- Intel® Xeon® Processor 7400 Series
- TPoX Benchmark
- Performance Results
  - Mixed Workload: 70% Query, 30% IUD
  - Insert Performance
- Summary

Benchmark Results vs. Customer requirements

- Financial Application Logging
  - Requirement: insert & index up to 50M docs per day, plus queries
  - Our benchmark exceeds this requirement
    - inserts 18M documents of similar size (4k to 20kb) per hour
    - processes 48M mixed transactions in 2 hours

- Non-US Government Agency
  - Requirement: store & index up to 1B *small* XML documents per day
  - Our benchmark:
    - inserts 714,285 order messages (1 – 2kb) per minute
    - projection: → 42.8M inserts per hour → 1B inserts in 24 hours
    - meets this requirement

Summary

- Can store & index 1TB of raw XML data in less than 500GB disk space
- It's not hard to tune XML-based transaction processing for good performance
- DB2's autonomic and self-tuning features do (most of) the job!
- Prerequisite: well-balanced hardware,
  CPU cores ↔ #disks ↔ main memory
- 50% more cores provide 48% more throughput,
  when upgrading from Intel® Xeon® 7300 processor series to Intel® Xeon® 7400 processor series
- DB2 + Linux + Intel®: A powerful combination to deliver high performance and scalability for XML data management workloads

Acknowledgement

- Thanks to the following people for their help and support with this benchmark:
  - Tuan Bui, Kshitij Doshi, Garrett Drysdale, Joe Ellis, AG Ramesh, Ying Zhang, Paul Gryskiewicz, Stephen P. Smith (Intel)
  - Tim Kiefer, Qi Jin (IBM Silicon Valley Lab)
  - Kevin Xie, Peter Shum, Berni Schiefer (IBM Toronto Lab)

Notices

- Copyright © 2008, Intel Corporation. All rights reserved.
- *Other names and brands may be claimed as the property of others.
- Intel® Xeon® Processor 7300 Series and Intel® Xeon® Processor 7400 Series are trademarks of Intel Corporation in the U.S. and other countries.
- Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit Intel Performance Benchmark Limitations

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, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.