TPCTC - Overview of TPC Benchmark E

Author: Trish Hogan
Presenter: Karl Huppler

Overview

- What is a benchmark?
- Why a new benchmark?
- What is TPC-E?
- Current Landscape

What is a benchmark?

Generic Computer Benchmark

- What is a benchmark?
- A defined workload that:-
  - Stresses a particular part of a system in a measurable way
  - Has a result/score you can use for comparison
  - Is repeatable
  - Is independently verifiable
  - Can be run by anyone
- Examples – TPC-E, TPC-C, SPEC CPU2006, SAP S&D

What is the TPC?

- Transaction Processing Performance Council
- The TPC is a non-profit corporation founded to define transaction processing and database benchmarks and to disseminate objective, verifiable TPC performance data to the industry.
- The TPC is a vendor-neutral and database-agnostic
- Uses Third Party Auditors to verify results

What is a benchmark?

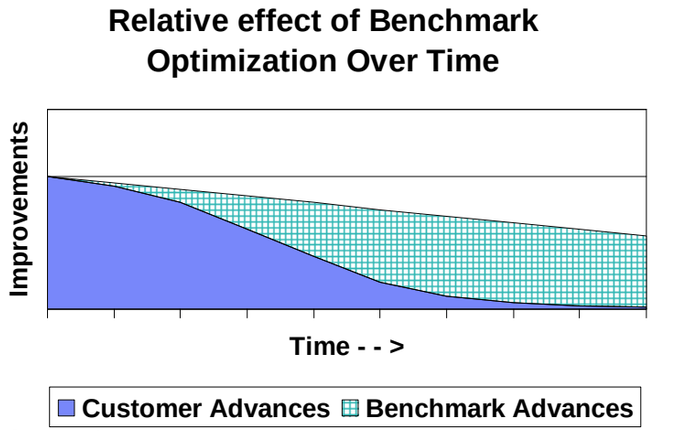
Why run benchmarks?

- Marketing
  - To differentiate (“Only me!”)
  - To include (“Me too!”)
  - Product launch
- Necessity
  - Government contract bids
  - Pricing practices
- Information
  - New products
  - Competition comparison
  - Sizing data

How customers use benchmarks

- Highlights Hardware and Software Performance
- Demonstrates serious enterprise worthy solution
- Choosing Servers
- Choosing Database Management Systems
- Capacity Planning

Why a new benchmark?

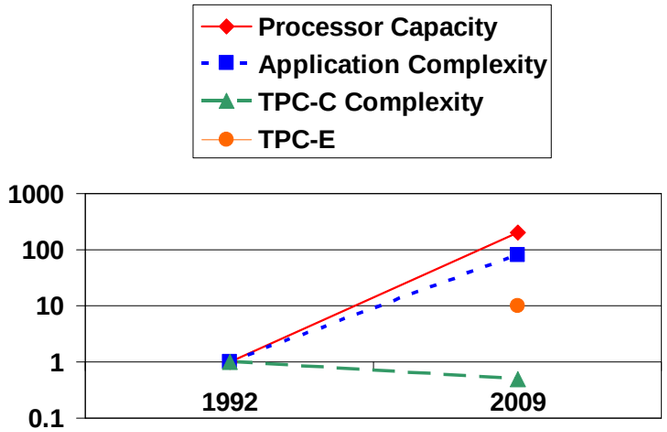


(concept chart not backed by statistics)

- Benchmarks have a life time
  - Good benchmarks drive industry and technology forward
  - At some point, all reasonable advances have been made
  - It is counterproductive to encourage artificial optimizations
  - So, even good benchmarks become obsolete over time
- TPC-C Specification approved July 23, 1992
  - It became de facto industry standard OLTP benchmark, but…
  - TPC-C is over 17 years old
    - In “dog years” that’s 119
    - In “computer years” it’s basically ancient!

Why a new benchmark?

Signs of TPC-C's Age



- Moore's Law (various versions)
  - The number of transistors in an electrical component double every two years
  - The capacity of a processor to do work doubles every 18 months

(concept chart not backed by statistics)

- The Java_C++_Middleware_x-generation-development-tools corollary to Moore's Law:
  - Application and data complexity will grow in direct relationship to the amount of compute capacity available to absorb it
  - double every 18 months? Maybe not, but significant
- TPC-C application complexity
  - Logically constant since 1992
  - Actually much reduced by 3-tiered configurations, database optimization and very talented application programmers

Why a new benchmark?

Signs of TPC-C's Age

- As a result
  - Larger and larger I/O subsystems needed
  - Benchmarks are too expensive
  - Benchmarks take too much time to run and audit
  - Workload and configuration are less representative to customer environment than when benchmark was first introduced

- Need a new benchmark – TPC-E approved February 2007

What is TPC-E?

TPC-E Benchmark

- An industry standard database server benchmark
  - Audited by a third party
  - Backed by the strength of the TPC
- Measures online transaction processing performance
  - Business critical applications
- Based on a viable business model
  - Emulates a brokerage house receiving trade requests, account queries and financial market queries

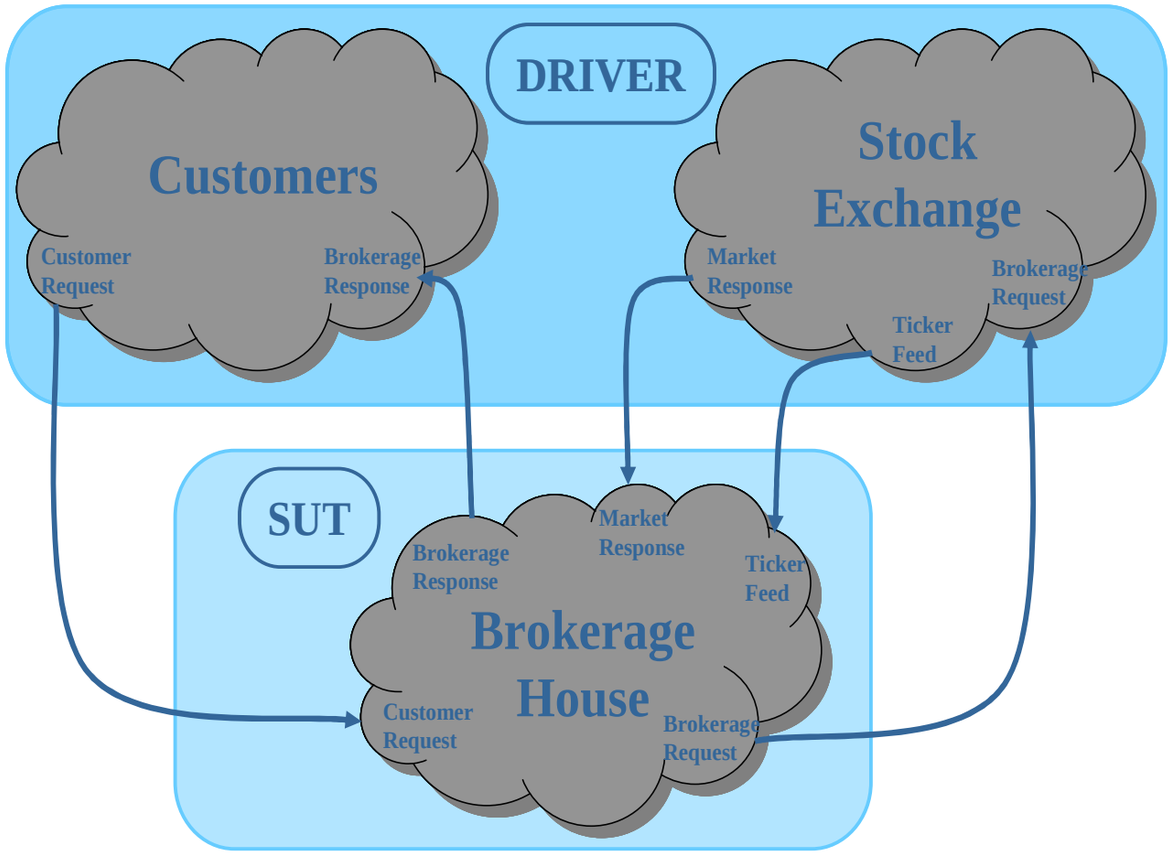
What is TPC-E?

TPC-E Benchmark

- Provides comparable results
  - i.e., results from different vendors can be compared
- Relatively reduced cost and complexity of running the benchmark
- Database schema reflective of a modern OLTP environment
- Encourages database optimizations that customers need

What is TPC-E?

Financial Model



What is TPC-E?

Financial Model

- TPC chose Financial model
  - Met volume/scaling and market relevance criteria
  - Interesting transaction profiles
  - Constant transaction profiles with regard to performance

What is TPC-E?

Tables

Customer	Broker	Market	
ACCOUNT_PERMISSION	BROKER	COMPANY	SECTOR
CUSTOMER	CASH_TRANSACTION	COMPANY_COMPETITOR	SECURITY
CUSTOMER_ACCOUNT	CHARGE	DAILY_MARKET	
CUSTOMER_TAXRATE	COMMISSION_RATE	EXCHANGE	
HOLDING	SETTLEMENT	FINANCIAL	Dimension
HOLDING_HISTORY	TRADE	INDUSTRY	ADDRESS
HOLDING_SUMMARY	TRADE_HISTORY	LAST_TRADE	STATUS_TYPE
WATCH_ITEM	TRADE_REQUEST	NEWS_ITEM	TAXRATE
WATCH_LIST	TRADE_TYPE	NEWS_XREF	ZIP_CODE

What is TPC-E?

Database Content

- Populated with pseudo-real data
- Distributions based on:
  - 2000 U.S. and Canada census data (*)
    - Used for generating name, address, gender, etc.
    - Introduces natural data skew
  - Actual listings on the NYSE and NASDAQ
- Benefits
  - Realistic looking data
  - Compressible for backup testing, etc.
  - Closer match to actual customer databases
  - Anticipate usage well beyond benchmark

What is TPC-E? – Database Content

- Sample data from TPC-E CUSTOMER table

C_TAX_ID	C_L_NAME	C_F_NAME	C_M_NAME	C_GNDR	C_DOB	C_EMAIL_1
757FI2006HD923	Mexicano	Courtney	T	F	1997-11-30	CMexicano@hotmail.com
922SN3775RQ823	Udley	Judith	F	F	1954-09-27	JUdley@earthlink.com
006GT3444BE624	Buchanan	John	R	M	1971-06-13	JBuchanan@msn.com
181UZ4114LR434	Soloman	Clinton	D	M	1938-02-27	CSoloman@rr.com
355IE4773VF335	Orner	Harry	P	M	1974-11-15	Horner@attbi.com

- Sample data from TPC-C CUSTOMER table

C_FIRST	C_MIDDLE	C_LAST	C_STREET_1	C_CITY
RONpTGcv5ZBZO8Q	OE	BARBARABLE	bR7QLfDBhZPHlyDXs	OmWlmelzIJ0GeP kYM
e8u6FMxFLtt6p Q	OE	BARBARPRI	eEbgKxoIzx99ZTD S	4V1t1VmdVcXyoTOMwpPz
bTUkSuVQGdXLjGe	OE	BARBARPRES	QCGLjWnsqSQPN D S	jVHBwIGFh2k oTOMwpPz
18AEf3ObueKvubUX	OE	BARBARESE	JnBSg4RtZbALYu S	5g8XMnlegn oTOMwpPz
mFFsJYeYE6AR bUX	OE	BARBARANTI	MLEwwdy3dXfqngFcE	yVVR4iEtj0ADEwe wpPz

What is TPC-E?

Transactions

Name	Symbol	Access	Description
Broker-Volume	BV	RO	DSS-type medium query
Customer-Position	CP	RO	"What am I worth?"
Market-Feed	MF	RW	Processing of Stock Ticker
Market-Watch	MW	RO	"What's the market doing?"
Security-Detail	SD	RO	Details about a security
Trade-Lookup	TL	RO	Look up historical trade info
Trade-Order	TO	RW	Enter a stock trade
Trade-Result	TR	RW	Completion of a stock trade
Trade-Status	TS	RO	Check status of trade order
Trade-Update	TU	RW	Correct historical trade info

What is TPC-E?

Performance Metric

- The metric for TPC-E is tpsE and $/tpsE
  - Transactions per second (not per minute)
  - The Trade-Result is the transaction that is counted
  - Trade-Result is only 10% of the transaction mix
  - Other transactions have consistent mix
    - Number Trade-Result transactions close to proportional to number total transactions
- tpsE pronounced “tipsy”

TPC-E Compared to TPC-C

Characteristics	TPC-C	TPC-E
Business model	Wholesale supplier	Brokerage house
Number of database tables	9	33
Number of database columns	92	188
Minimum columns per table	3	2
Maximum columns per table	21	24
Datatype count	4	Many
Primary keys	8	33
Foreign keys	9	50
Tables with foreign keys	7	27
Check constraints	0	22
Referential integrity	No	Yes
Database roundtrips per transaction	One	One or many
Number of transactions	5	10
Number of physical I/Os	5x	x
RAID requirements	Database log only	Everything
Timed database recovery	No	Yes
TPC-provided code	No	Yes

What is TPC-E?

An improvement over TPC-C

- Benchmark configuration closer to customer configurations
  - Makes it easier to use benchmark results for capacity planning
  - Drives hardware and software design that benefits customers
    - Data accessibility tests – drive storage technology improvements
    - Database Recovery Time – drive DB recovery improvements
    - Isolation tests – drive DB locking improvements
    - Secondary index, constraint checking, database loading
- Less expensive for hardware vendors to run than TPC-C
  - So vendors will be able to publish more results
  - Customers will have more comparison information

What is TPC-E?

An improvement over TPC-C – less TPC-E hardware needed

	TPC-E	TPC-C
System	IBM System x3650 M2	HP ProLiant DL370 G6
Total price	$260,792	$678,231
Score	817.15 tpsE	631,766 tpmC
Price/performance	$319.15/tpsE	$1.08/tpmC
Availability Date	7/31/09	3/30/09
Processors/cores/threads	2/8/16	2/8/16
Database server memory	96GB	144GB
Disks	392 (RAID-10)	1,210 (RAID 0)
RTEs	0 (N/A)	18
Clients	2	8
Total systems to setup & tune	3	27

Comparison of TPC-C and TPC-E configurations for two-processor server

Current Landscape

- TPC-E momentum increasing
  - 27 TPC-E results
  - From 6 different vendors
  - System ranging from 1 to 16 processors
  - As of 18 August, 2009 TPC-E publishes outnumber 2009 TPC-C publishes
  - To date, all publishes use Microsoft SQL Server

- Future
  - Need other database vendors to publish TPC-E results
  - Expect TPC-E popularity to improve if TPC delivers an Energy metric in addition to the current performance, price/performance and availability metrics

Summary

TPC-E is a new Database Server Benchmark

- Use TPC-E Results to
  - Get customer's attention
  - Differentiate Hardware/Software/Solution
  - Launch New Hardware/Software
  - Help with capacity planning
- TPC-E Improvements
  - Drives more relevant product improvements
  - More complex database schema and transactions
  - Reduced physical I/Os per second
  - Reduced benchmark cost
- TPC-E popularity is gradually growing

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States, other countries, or both.

Not all common law marks used by IBM are listed on this page. Failure of a mark to appear does not mean that IBM does not use the mark nor does it mean that the product is not actively marketed or is not significant within its relevant market.

Those trademarks followed by ® are registered trademarks of IBM in the United States; all others are trademarks or common law marks of IBM in the United States.

For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml:

*, AS/400®, e business(logo)®, DBE, ESCO, eServer, FICON, IBM®, IBM (logo)®, iSeries®, MVS, OS/390®, pSeries®, RS/6000®, S/30, VM/ESA®, VSE/ESA, WebSphere®, xSeries®, z/OS®, zSeries®, z/VM®, System i, System i5, System p, System p5, System x, System z, System z9®, BladeCenter®

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
TPC, TPC Benchmark, TPC-E and TPC-C are trademarks of the Transaction Processing Performance Council.
Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
Microsoft, Windows, SQL Server, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
UNIX is a registered trademark of The Open Group in the United States and other countries.
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

* All other products may be trademarks or registered trademarks of their respective companies.

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.
IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.
All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.
Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.