

THE HISTORY OF DEBITCREDIT AND THE TPC

by Omri Serlin

This is a personal account of how the Transaction Processing Performance Council (TPC) came into being and how it created TPC Benchmark(tm) A and B (commonly abbreviated TPC-A, TPC-B). I also describe how these tests differ from each other, and from the DebitCredit and TP1 tests, their popular but non-standardized predecessors. Finally, I offer an assessment of the value of the standards developed by the TPC, and how they compare with standards developed by other recently-formed performance measurement standards bodies, most notably SPEC (System Performance Evaluation Cooperative).

Interest in OLTP Performance

In the early 1980s there was a renaissance of interest in characterizing the performance of on line transaction processing (OLTP) systems. This interest was driven by two seemingly contradictory trends. First, there were expectations of an increased demand for very high performance transaction systems. This demand was expected to arise from the increasing automation of common, daily business transactions. The proliferation of automatic teller machines (ATMs) in the banking industry was often cited as a prime example of the trend. Indeed some banks were busily developing large systems to handle the anticipated deluge of transactions arising from, for instance, gasoline stations converting their pumps to automatic operation, controlled by debit or credit cards. Interestingly, the deluge did not materialize, although card-operated fuel pumps have become rather popular in Europe.

In any event, the expectation of sharply increased transaction loads gave rise to intense debates among workers in the field about "1K tps" systems, i.e., those capable of sustaining 1000 transactions per second (tps). IBM publicized the results of an OLTP test, known as the "1K" test, which, as its name implied, was meant to prove specifically that an IBM mainframe could indeed perform 1,000 tps (REF: FTSN-63). There was even a special conference launched to deal with this issue; dubbed HPTS, the International Workshop on High Performance Transaction Systems is a bi-annual conference which has been held three times since 1985.

At the other end of the spectrum, the success of Tandem Computers created a good deal of interest in medium-range OLTP systems. About a half-dozen would-be Tandem competitors

entered the market in the early 1980s, expecting to derive their chief competitive edge from the use of high-performance microprocessor technology, compared with the minicomputer-style architecture and ad-hoc logic employed in the Tandem systems. With the notable exception of the highly-successful Stratus Computer, most of the "new wave" OLTP vendors soon fell by the wayside. Nevertheless, while they were actively making performance claims, the issue of OLTP performance characterization remained red-hot.

While there was general agreement that such metrics as mips and Dhrystones were wholly inadequate to characterize the performance of OLTP systems, there was no agreement on exactly how tps ratings were to be derived. Each vendor selected a test to suit that vendor's specific market orientation and system idiosyncracies. Furthermore, in making their tps claims the various vendors released few details regarding the test. Thus while plenty of tps claims were being unleashed, it wasn't at all clear whether any of the published tps ratings were in any sense comparable (REF: FTSN-47a).

"Anon et. al." Publishes a Paper

Against this background, in the summer of 1984 Jim Gray of Tandem Computers wrote an early version of what eventually became the celebrated "Anon et.al." paper. This early version was distributed by Jim to a number of interested Tandem employees, as well as to nineteen other professional acquaintances in industry and academia for comments and suggestions. "Anon et. al." was graciously used by Gray to suggest that the paper was authored by this entire group.

The paper recommended the adoption of three standard performance tests for OLTP systems: one on-line transaction processing test, dubbed DebitCredit; and two batch tests, a sort and a scan. The rationale for these three tests was that they would adequately characterize the key aspects of a system intended for commercial on-line transaction processing work.

For the on-line test, the paper proposed a highly stylized emulation of a teller support system in a large, multi-branch bank. The key parameters of the emulated bank (e.g., number of customer accounts, number of branches, number of tellers per branch) roughly corresponded to the actual state of a well-known California bank (Bank of America) in the early 1970s. Indeed this emulation was carried out by IBM, where Gray worked at that time, in order to analyze the performance bottlenecks and price-performance considerations that led the bank to select a minicomputer-based architecture for its initial teller support system, in preference to IBM mainframes. Gray's paper named this test "DebitCredit", although it also suggested the names

"ET1" and "TP1", which were internal IBM code names. As it turned out, this profusion of terms led to very undesirable consequences later.

Although Gray's paper described the DebitCredit test (and the other two tests) in broad functional terms only, it did already contain most of the key ideas later codified in TPC-A and TPC-B. These could be summarized as simplicity, applicability to distributed systems, scalability, comparability, and vendor neutrality.

Key Characteristics of DebitCredit

For simplicity, only a single transaction type was included, representing a deposit by an account holder. The deposit activity was to be recorded in three randomly-accessed indexed files: the account file, the branch file, and the teller file. In addition, the transaction details were to be recorded in a sequential history file. These four files are often abbreviated the ABTH files.

A measure of application realism was included by requiring that 15% of all incoming transactions involve account numbers residing at a different branch than the one at which the transaction was processed. This represented customers doing business at other than their "home" branches. More importantly, it was hoped that in systems employing multiple loosely-coupled processors (such as those from Tandem and Stratus), this provision would force some inter-CPU communications and possibly invoke distributed transaction integrity mechanisms.

To assure scalability, the rate at which each teller's terminal generated transactions was fixed (100 seconds per transaction in the Gray paper); and the size of the account, branch, teller, and history files was to be a function of the system's throughput; specifically, the Gray paper suggested a scaling of, for each input tps, 100,000 account records, 10 branches, and 100 tellers. In addition, the sequential History file was meant to hold a record for each transaction over a 90 day period, although the Gray paper did not specify whether the history data was to represent 24-hour-a-day or 8-hour-a-day operation.

The tps throughput of an OLTP system is closely dependent on the allowable response time constraints. When longer response times are permitted, higher tps ratings can generally be reported. In order to make tps ratings comparable, the paper proposed a constraint under which 95% of all transactions completed in 1 second or less.

Another important aspect of comparability is configuration control. The usual industry practice up to that time was to compare configurations that were in some sense equivalent (having the same amount of memory, disks, etc.) and report some CPU-only performance metric along

with purchase costs of the systems being compared. The Gray paper introduced a novel idea in this respect. Instead of configuring equivalent systems, Gray suggested that the cost of the tested system be used as the rationalization factor. Thus vendors would be free to configure any system they deemed appropriate; but, in addition to the tps ratings, they would be required to publish the cost-per-tps for the tested system. In addition, Gray suggested that the costs in question should include not merely the initial purchase price, but also the costs of software licenses and maintenance for a period of 60 months, the so-called "5-year-cost-of-ownership".

The Gray proposal was "vendor-neutral" in two important respects. First, by stating the requirements of the test in a widely-available trade publication, the test was, in effect, put in the public domain. Furthermore, by specifying the requirements at a high functional level, rather than as an executable program, the test became implementable on any hardware platform using any desired software components. No other OLTP test enjoyed these advantages.

Early Efforts to Garner Industry Consensus

When I received Gray's original paper, I was already engaged in a modest effort, through my FT Systems newsletter as well through personal contacts, to form an industry-consensus forum to establish performance measurement standards in the OLTP field. Dubbed the Working Group on Performance Measurement Standards (WG-PMS), this was a very loose association of (at its peak) some 50 individuals who in various forms had expressed interest in the subject.

Recognizing the relevance of the paper's suggestions, I immediately dashed off copies to the WG-PMS, eliciting reactions or counter-proposals. To my great surprise, no one reacted. The lack of reaction convinced me that a substantial effort would be required to make the organization into a going concern; an effort which at the time I was not prepared to make. To my recollection, WG-PMS activity ceased altogether sometime in 1985.

Gray, however, continued work on his paper, which was published as Tandem Technical Report 85.2 in February, 1985. A slightly-edited version of that report appeared under the title "A Measure of Transaction Processing Power" by Anon et.al. in the April 1, 1985 issue of Datamation magazine (REF: ANON). The actual names of the 23 members of the "et. al." group were listed in TR 85.2 but not in the Datamation article.

The 1985 Datamation article proposed a rough guide for performance under the proposed benchmark, or reasonable facsimiles thereof (TABLE 1). These results were derived or extrapolated from some actual tests, but the systems tested were not explicitly identified.

Table 1. "TYPICAL" DebitCredit Results

Rating	K-Inst.	I/O	tps	\$K/tps	Packets
Lean & Mean	20	6	400	40	2
Fast	50	4	100	60	2
Good	100	10	50	80	2
Common	300	20	15	150	4
Funny	1000	20	1	400	8

K-Inst. is the number of instructions (in thousands) executed on behalf of the transaction. I/O is the number of actual disk I/Os per transactions. tps is transactions per seconds, assuming the 95%-under-1-second response. \$K/tps is 5-year-cost-of-ownership divided by the tps rating. Packets refers to X.25 packets (the Gray proposal assumed an X.25 network linking the teller terminals to the central system).

While the two proposed batch tests, sort and scan, were largely ignored, the proposed on-line test quickly gain wide popularity following the Datamation article, with many vendors issuing a variety of "ET1" and "TP1" claims.

However, rather than decreasing the confusion, the publication of the article merely encouraged vendors to claim performance under the "industry standard benchmark". It soon became quite clear that the article was not sufficiently precise to allow unambiguous implementations (REF: FTSN-47b, -47c). Furthermore, even those aspects that were quite clearly stated were often violated. As an example, few of the published results used the proposed 5-year-cost-of-ownership formula, preferring instead to use initial hardware costs only.

TP1 Variation Muddies the Waters

The situation became even more critical around 1987, when suppliers of relational database systems began publishing so-called "TP1" results with tps ratings that appeared to be abnormally high relative to the power and configuration of the hardware platforms employed in these tests. By early 1988, it became clear that while these "TP1" tests did use some variation of the DebitCredit transaction, they included a major simplification by ignoring entirely the user terminals and the X.25 connecting network (REF: FTSN-67).

Instead, TP1 as used by the database vendors relied on batch-type "transaction generator" processes that created transactions as fast as possible, without allowing for any "think time". In practical terms, this TP1 variation assumed that the transactions had already arrived at the

memory of the system under test (SUT). This was essentially true regardless of whether the transaction generators executed on the SUT itself, or were placed on an external system connected to the SUT via a local area network (LAN).

In either case, the SUT was not burdened by the management of the connecting network and the large number of terminals that would have been required for the reported tps rate under a full DebitCredit implementation. The rationale for this simplification was that the database vendors were interested in characterizing the performance of their product while avoiding as much as possible the impact of extraneous factors, such as the efficiency of the operating system in handling terminal communications.

While this simplification made good sense from the point of view of the database vendors, it introduced a significant new uncertainty regarding the comparability of results from the so-called "industry standard" OLTP test. In REF: FTSN-67, which described the difference between the original DebitCredit and the new TP1 tests, I suggested that vendors should report the former as "Class I" results and the latter as "Class II" results.

There was no industry response to my proposed classification scheme; stronger action was called for if further confusion regarding OLTP performance claims was to be avoided.

The Sawyer-Serlin Paper

By that time the practice of hiring an auditor to certify the results of an OLTP test was beginning to take root. This trend began in March, 1987 when Tandem hired Codd & Date Consulting to certify the 208 tps result obtained under a version of DebitCredit on a 32-processor Tandem VLX system, using the NonStop SQL relational database system (REF: FTSN-55). Tandem eventually published and made available to the public at large massive documentation supporting the performance and price performance claims, and giving details of important parameters and configuration factors with potential significant impact on the performance of the tested system (REF: TNDM). This document later became the model for what the TPC codified as the "Full Disclosure" report, a concept I initially proposed as the "Reporting Checklist" in REF: FTSN-47c.

Tom Sawyer of Codd & Date Consulting was the auditor on the Tandem NonStop SQL test; subsequently he officially witnessed a number of the database-only, TP1-type tests. Through this work, Sawyer became acutely aware of the ambiguities in the Anon et. al. article, and the confusion between the original DebitCredit and the newer TP1 tests. He was therefore very interested in establishing a more rigorous standard to help him in auditing both types of tests.

Thus when I approached him in June, 1988 regarding the possibility of generating a proposed standard document, he responded enthusiastically.

Sawyer and I spent a few evenings hammering out a proposed standard document for the DebitCredit and TP1 tests. The document, dated June 22, 1988, was titled "DebitCredit Benchmark - Minimum Requirements and Compliance List" and comprised of 13 pages, including several pages of comments and a reference list. The paper was included in its entirety in REF: IBM.

In this paper we devised a scheme which we hoped would permit a rational combination of both the simplified TP1 test and the full-scale DebitCredit benchmark. The scheme consisted of a series of mandatory requirements, roughly corresponding to a TP1-type test. A test sponsor would earn 70 points for meeting the minimum requirements (essentially a TP1 test); 20 additional points for proper terminal network representation (essentially a DebitCredit test); and 10 additional points for protecting the transaction log from single failures (a higher measure of fault-tolerance). Our initial idea was that test sponsors would report the point rating in addition to the tps throughput and cost-per-tps figures.

Later, other schemes were proposed, such as reducing the attained tps throughput by multiplying it by the attained point percentage rating. All of these ideas were eventually rejected by the TPC, although in October, 1988, IBM actually managed to conduct a test on several 9370 and 4381 models using the proposed Sawyer-Serlin document as its authority (REF: IBM).

It is worth noting that this test was awarded a score of 85 by the auditor, Tom Sawyer, meaning that it was a pretty complete DebitCredit test. Yet a similar DebitCredit test run by DEC on at least one virtually identical IBM model yielded substantially poorer results. This discrepancy provided additional fuel to the on-going benchmarking wars (FTSN-75).

The TPC Gets Launched

Armed with the Sawyer-Serlin paper I then launched a new drive to form an industry-consensus body for creating and enforcing OLTP performance measurement standards. This time, the effort was successful. By the 10th of August, 1988, eight companies had agreed to join the newly-formed Transaction Processing Performance Council (TPC), and an announcement to that effect was sent over the Business Wire on that day. The initial members were Control Data Corp.; Digital Equipment Corp.; ICL; Pyramid Technology; Stratus Computer; Sybase; Tandem Computers; and Wang Laboratories. By the end of 1988, there were 26 member companies. At this writing, the membership list stands at 35, including all the major U.S. system and database

vendors, and a strong international contingent with both European and Far East representation (the complete list is in TABLE 2).

Table 2. TPC membership as of August 1989.

Arix	Informix	Sequoia Systems
AT&T	Ingres	Siemens A.G.
Bull	ITOM Int'l Co.	Software A.G.
Computer Associates	Mitsubishi Electric	Stratus Computer
Control Data Corp.	NCR	Sun Microsystems
Data General	NEC Corp.	Sybase
Digital Equipment Corp.	Oki Electric	Tandem Computers
Fujitsu America	Olivetti	Teradata
Hewlett Packard	Oracle	Unify
Hitachi Ltd.	Prime Computer	Unisys
IBM	Pyramid Technology	Wang Laboratories
ICL	Sequent Computer	

Originally I expected (somewhat naively) that the Council would achieve all of its work by means of audio teleconferences and written communications. However, it soon became quite clear that the bandwidth of audio teleconferences was too limited for the type of detailed discussions required, and written communications too slow. Thus face-to-face meetings were indicated. Indeed, the TPC eventually required seven General Meetings and 11 teleconferences to hammer out its first standard, TPC Benchmark(tm) A, a 42-page document which was published in November of 1989.

Early in its deliberations, the Council rejected the previously-discussed idea of combining the DebitCredit and TP1 tests and opted instead for two separate standards. TPC-A (REF: TPCA) is the Council's version of the DebitCredit test, while TPC-B, which was officially approved in August, 1990 (REF: TPCB), represents the Council's version of the TP1 test. Accounts of the progress of the TPC were given in SERL1 and SERL2.

TPC-A differs from DebitCredit in several important respects. These are summarized in TABLE 3.

TABLE 3: TPC-A Cvcompared to DebitCredit.

Item	DebitCredit	TPC-A
Authority	None - published as a trade press article computer performance	Approved by TPC, the largest industry body dealing with
Transaction Profile	Terminal I/O within transaction	Terminal I/O outside transaction New Account Balance returned to terminal
Profile	transaction	transaction
System Properties	Duplexed log	Strong ACID with tests specified.
Response Time	95% under 1 second; measured at SUT.	90% under 2 seconds; measured at driver.
File Scaling (per tps)		
- Branch records	10	1
- Teller records	100	100
- Account records	100,000	100,000
- History records	unclear	2,592,000
Transaction Arrival	Not Specified	Random (Negative Exponential Distribution)
Throughput Calculation	Only one tps point required	Complete tps vs. RT graphs required.
Number of Terminals	100 per tps.	10 per tps.
Terminal Type	3270-type only.	Intelligent "client" workstations as well as "dumb" terminals.
History File	Unified	Horizontal partitioning permitted.
Interconnect Method	Wide-area X.25 network	WAN or LAN with any standard protocol.
Required Storage	90 days of history data.	8 hours of actual SUT operation; plus 90 days history to be priced.
Pricing	Only "computer room" equipment & software over 5 years.	All hardware and software over 5 years except physical comm. media.
Full Disclosure	No mention.	Detailed report required; auditing recommended.
Reporting Metrics	tps and cost-per tps.	tpsA-Wide and/or tpsA-Local and cost- per-tps.
LEGEND:		
ACID - Atomicity, Consistency, Isolation, Durability	SUT - System Under Test	
RT - Response Time	WAN, LAN - Wide (Local) Area Network	
tps - Transactions per Second.		

By far the chief difficulty the TPC faced was that the DebitCredit test as originally formulated assumed a model in which a "glass-house" mainframe served relatively low-intelligence teller terminals over a wide area network. In contrast, the majority of TPC members were interested in smaller systems as well, especially ones that employed local area networks and/or the "client-server" model, with intelligent workstations taking the place of teller terminals. In an attempt to reach a compromise on these and other issues, the TPC clearly relaxed the "rules of the game" in a number of important areas in defining TPC-A.

For example, while the original benchmark demanded that 95% of all transactions must meet a 1-second response time criterion, TPC-A is satisfied with 90% of the transactions responding in 2 seconds. All else being equal, this means TPC-A results in substantially higher tps ratings.

Another significant easing of the requirements is in the area of transaction arrival time, or "think time". TPC-A can be satisfied with a 10-seconds average think time, meaning 10 terminals per tps; the original benchmark specified 100 seconds, meaning that 100 terminals must be emulated for each tps achieved. Again, this relaxation leads to higher tps ratings.

The removal of the requirement that there be only one history file in the system under test (SUT) removes the "hot spot" that is created as all transactions attempt to append a record at the end of that sequential file. This, too, improves tps throughput.

Finally, whereas DebitCredit visualized a configuration in which all terminals connected to the SUT over a wide-area, X.25 communications network, TPC-A also permits a local-area interconnect. Since it is generally believed that the LAN interconnect results in higher throughput, TPC-A demands that, in order to make clear which interconnect scheme was used, throughput must be stated as either tpsA-Wide or tpsA-Local, not merely tps.

The "A" appended to the term "tps" is meant to signify that the results are derived from TPC-A and not from some other OLTP test. Since OLTP tests by their very nature are highly non-linear, the TPC standard specifically forbids test sponsors from making comparisons between TPC-A results and any other TPC or non-TPC benchmark.

While it could be argued that the simplifications described above, and others, further weaken the ability of the test to represent a real-life application, it is important to remember that the principal consideration in each case was the attainment of a consistent yardstick that could be applied to widely-diverse systems. Application realism took second place wherever it conflicted with the common yardstick goal.

It is also important to note that TPC-A introduces more severe requirements in several important areas. For example, response time is to be measured at the driver system, where the terminals are emulated, rather than at the SUT. Thus the delays across the communications network are included in the RT, somewhat depressing the potential throughput. This turned out to be one of the happy instances where application realism (in this case, the measurement of RT as the terminal user would see it) also coincided with practical considerations (it is easier to measure RT at the driver).

Random rather than uniform transaction arrival rate is another example of a stiffened requirement. The Anon et.al. paper didn't specify this point; there is some theoretical and practical evidence that allowing uniform arrivals doubles the transaction rate.

While DebitCredit only required the log file to be duplexed, allowing the system to maintain transaction integrity in the presence of a single failure, the TPC-A standard explicitly requires the SUT to meet fairly-severe ACID (Atomicity, Consistency, Isolation, and Durability) properties; it even specifies tests to establish the presence of these properties.

By requiring a Full Disclosure Report, TPC-A makes a critically-important contribution to "truth-in-advertising". Those who wish to make claims under the TPC standard must submit to the TPC, as well as make available to the public at reasonable cost, a rather detailed report which gives many parameters and further specifies the test environment. Today, these parameters and environmental factors, which can be crucial to the full understanding of the test results, often go unreported.

TPC-A Official Results

As of this writing, three companies (DEC, HP, and IBM) have reported results under TPC-A. TABLE 4 summarizes the published results. It is important to consult the relevant Full Disclosure reports for additional information and qualifications of these results.

Table 4. Reported TPC-A results.¹

System	tpsA-Local	K\$/tps
VAX 9000-210 w/DSM V6.0	143.4	29.3
VAX 9000-210 w/Rdb 3.1	69.4	40.0
VAX 4000-300 w/DSM V6.0	41.4	24.1
HP Series 960	38.2	32.6
HP Series 949	32.2	26.9
IBM AS/400 Model B70	27.1	31.3
VAX 4000-300 w/Rdb 3.1	21.7	31.9
VAX 4000-300 w/Rdb 4.0	21.6	32.1
HP Series 932*	13.6	33.1
IBM AS/400 Model B45	8.98	37.0
IBM AS/400 Model C25	7.75	29.7
HP Series 922	7.7	33.1
HP Series 920*	4.95	33.0
* Unaudited.		

All of the test sponsors to date have chosen the local interconnect option, and hence have reported tpsA-Local numbers. The cost-per-tps is generally better than Gray's original "lean-and-mean" figure (\$40K/tps, see Table 1) partly because of advances in hardware and software, and partly because of the more relaxed TPC-A think time and response time constraints. In analyzing the cost-per-tps figures it is important to remember that the priced configuration may include one or more front-end processors, in addition to the main system indicated in the table. The Full Disclosure reports provide detailed configuration data.

¹Copies of the Full Disclosure reports may be obtained from the following addresses:

Administrator, TPC Benchmark Reports TPC Benchmark A Administrator
Transaction Processing Systems Group IBM Performance Evaluation Center, 40-F3-01
Digital Equipment Corp. 1 East Kirkwood Blvd.
151 Taylor Street Roanoke, TX 76299-0015
Littleton, MA 01460-1407

Commercial Systems Division
Attention: Colleen Doyle M/S 44MF
Hewlett Packard Company
19111 Pruneridge Avenue
Cupertino, CA 95014

General TPC information and published standard specifications may be obtained from the TPC Administrator at:

Waterside Associates
39510 Paseo Padre Parkway, Suite 350
Fremont, CA 94538
Phone (415) 792-2901
FAX (415) 792-4748
e-mail: shanley\@cup.portal.com

It is interesting to note that TPC-A results do not display the clear economies scale that the Anon et.al. paper discovered (Table 1). There is, in fact, some evidence that, all else being equal, systems that deliver higher tps throughput sometimes do so at a poorer cost-per-tps than systems with more modest tps ratings. This can be seen in the DEC and HP results, for instance. This dis-economy of scale tends to match comparisons based on cost-per-mips, where, for example, it is quite clear that an IBM PC provides a much better price-performance than an IBM 3090 mainframe.

TPC-B

TPC-B uses the same transaction profile, ACID requirements, and costing formula as TPC-A; but it permits the use of batch transaction generator processes, instead of terminal emulation, for creating incoming transactions. Because in this configuration the concept of a "user" doesn't really exist, the term "response time" has been replaced by "residence time", indicating essentially how long the transaction resided within the database server. Thus the horizontal axis of the two required performance curves is replaced in one case by residence time, and in the other by "level-of-concurrency". The level-of-concurrency is defined as TR , where T is the reported tpsB rate and R the average residence time.

Future Work

At this time, the TPC is already at work on two new benchmark proposals. One is derived from an inventory control and order entry application; it employs several transaction types of varying complexity (compared to just one transaction type in TPC-A and TPC-B), and may include concurrent batch work as well. Another proposed benchmark uses complex query-only transactions to emulate a decision support environment, using a relational database.

Assessment

I calculate that the TPC required nearly 1200 man-days of effort to define TPC-A. This figure excludes time spent internally in each company on critiquing and amending the intermediate drafts. When travel time and expenses are added, the resulting cost to the industry has been quite significant. It is fair to ask whether the results are worth all this effort.

The success of any standard has to be measured principally by the degree to which vendors and users adhere to that standard. In that respect, the TPC standards appear to be lagging those established by SPEC, the System Performance Evaluation Cooperative. Some dozen vendors have published results from the SPEC benchmark suite, which was launched and

completed at approximately the same times as TPC-A; despite a much larger TPC membership, only three vendors did the same for TPC-A during that time.

However, it is important to keep in mind the differences between the two cases. SPEC benchmarks continue the tradition of testing CPU performance exclusively; and, they are supplied as ready-to-run, executable programs. In contrast, TPC benchmarks measure the performance of the entire system, from user terminals to the back-end database and back. They are by nature far more complex to run, and very much more complex to document (due to the Full Disclosure requirements).

Furthermore, recall that the first two TPC benchmarks are stated as high-level specifications only, which is essential to assure that they are executable using any hardware platform, programming language, operating system, and database or file system. This advantage comes at a cost: each vendor must re-implement the benchmark from the ground up. That slows down the initial rate at which new vendors release TPC numbers.

It is also important to note that unlike SPEC results, TPC benchmarks are typically audited by an external auditor, an expensive and time-consuming process.

In terms of user acceptance, I am aware of a handful of RFQs to date that have specified TPC-A as a mandatory requirement. However, a number of queries from local, federal, and foreign governments, and some large end users, leads me to believe that the number of RFQs specifying TPC benchmarks will increase in the near future.

A lingering question is whether the TPC should establish a mechanism to endorse results from TPC benchmarks, perhaps in the same fashion that the American Dental Association gives seals of approvals to various tooth-pastes. The Council briefly considered this idea and tabled it after potential legal complications were uncovered. It is more likely that the TPC will opt to formally publish summaries of TPC results as stated in official full-disclosure reports.

It is already clear that some ambiguities and defects persist in the published TPC standards. The Council will no doubt at some point have to either establish a procedure for updating already-published standards (the Council already agreed that no standard would be changed within the first 12 months of publication); or to introduce new standards to replace older ones.

Even if TPC standards fail to elicit wider adherence, I believe that some of the key concepts introduced by the TPC will have lasting impact on future performance work. Among these is the Full Disclosure concept, namely the idea that, to be credible, a test sponsor must

furnish to the public sufficient back-up data to allow independent observers or end-users to judge the validity of the associated performance claims.

Independent auditing, while not a mandatory requirement, is highly recommended in the TPC standards. Despite the high cost, this practice has been accepted, in OLTP performance measurement, as the rule rather than the exception.

The idea that a whole-system test should be specified at a high- enough level to permit implementations on widely-divergent hardware and software configurations is of fundamental importance and may well become the norm in future OLTP performance work.

Concluding Caveat

Users are always better off if they can get vendors to test proposed systems under benchmarks that represent the user's actual workloads. Users should also be aware that results obtained under TPC-A or TPC-B cannot be used to predict an OLTP system's performance under drastically different workloads. Nevertheless, as comparative tools, the TPC-standardized benchmarks are more useful than nonstandard tests in the many cases where users do not wish to create and supervise custom benchmarks. Users who create custom benchmarks may also find the TPC standards useful as a model for stating the requirements for such benchmarks.

Copies of the Full Disclosure reports may be obtained from the following addresses:

Administrator, TPC Benchmark Reports
Transaction Processing Systems Group
Digital Equipment Corp.
151 Taylor Street
Littleton, MA 01460-1407

TPC Benchmark A Administrator
IBM Performance Evaluation Center, 40-F3-01
1 East Kirkwood Blvd.
Roanoke, TX 76299-0015

Commercial Systems Division
Attention: Colleen Doyle M/S 44MF
Hewlett Packard Company
19111 Pruneridge Avenue
Cupertino, CA 95014

General TPC information and published standard specifications may be obtained from the TPC Administrator at:

Waterside Associates
39510 Paseo Padre Parkway, Suite 350
Fremont, CA 94538
Phone (415) 792-2901

FAX (415) 792-4748
e-mail: shanley\@cup.portal.com

REFERENCES

- ANON Anon et.al., A Measure of Transaction Processing Power; Datamation, April 1, 1985, p. 112. Also Tandem Technical Report TR 85.2.
- FTSN-47a The TP1/ET1 Fiasco, FT Systems newsletter no. 47, July 1986, pp. 5-6. Discusses how the Stratus TP1 test differed from DebitCredit.
- FTSN-47b Editor's Notebook, FT Systems newsletter no. 47, July 1986, p.1. Decries the confusion in OLTP claims despite the Anon et.al. article.
- FTSN-47c DebitCredit: A Standard? FT Systems newsletter no. 47, July 1986, pp. 2-8. Points out ambiguities in the Anon et.al. article, suggests a "Reporting Checklist", and discusses some test results.
- FTSN-55 Tandem Performs Massive DebitCredit Benchmark, FT Systems newsletter no. 55, March 1987, pp. 4-8. Describes the Tandem NonStop SQL DebitCredit test on a 32-processor VLX system.
- FTSN-630 NEKAY: IBM's 1K tps Transaction Benchmark, FT Systems newsletter no. 63, November 1987, pp. 1-5. Describes the 1K benchmark and its results on a 3090-400 with FastPath.
- FTSN-67 TP1 vs. DebitCredit: What's in a Name? FT Systems newsletter no. 67, March 1988, pp. 1-7. Describes the differences between DebitCredit and the database-only TP1 variant; discusses and gives references to several TP1 reports.
- FTSN-75 IBM, DEC Disagree on DebitCredit Results, FT Systems newsletter no. 75, November, 1988, pp. 1-5. Analyzes the conflicting IBM and DEC claims for DebitCredit on an IBM system.
- IBM Auditor's Report of IBM Mid-Range DebitCredit Results Announced October 11, 1988. Also contains a complete reprint of the Sawyer-Serlin paper. Updated December 1, 1988. Published by IBM but has no publication no. or ordering information. A summary description can be found in FT Systems newsletter no. 75, November 1988, p. 14; REF FTSN-75; and FT Systems newsletter no. 77, January 1989, p. 10.
- SERL1 Serlin, O.: Toward an Equitable Benchmark, Datamation, February 1, 1989, p. 47. A TPC progress report.
- SERL2 Serlin, O.: Measuring OLTP With a Better Yardstick, Datamation, July 15, 1990, p. 62. Another TPC progress report.
- TNDM Tandem NonStop SQL Benchmark Workbook, Tandem pub. no. 84160, March, 1987. The first substantial Full Disclosure type report.

TPCA TPC Benchmark(tm) A Standard Specification, November 1989. In this volume.
TPCB TPC Benchmark(tm) B Standard Specification, August 1990. In this volume.