Shaping the Landscape of Industry Standard Benchmarks: Contributions of the Transaction Processing Performance Council (TPC)

Raghunath Nambiar[1], Nicholas Wakou[2], Andrew Masland[3], Peter Thawley[4], Matthew Lanken[5], Forrest Carman[6], and Michael Majdalany[7]

[1] Cisco Systems, Inc., 3800 Zanker Road, San Jose, CA 95134, USA
rnambiar@cisco.com
[2] Dell Inc., One Dell Way, Round Rock, TX 78682, USA
Nicholas_Wakou@dell.com
[3] NEC Corporation of America, 14335 NE 24th Street, Bellevue, WA 98007, USA
andy.masland@necam.com
[4] Sybase, an SAP Company, 1 Sybase Drive, Dublin, CA 94568, USA
peter.thawley@sybase.com
[5] Oracle Corporation, 500 Oracle Pkwy, Redwood Shores, CA 94065, USA
matthew.lanken@oracle.com
[6] Owen Media, 3130 E. Madison St., Suite 206, Seattle, WA 98112, USA
forrestc@owenmedia.com
[7] LoBue & Majdalany Magt Group, 572B Ruger St. San Francisco, CA 94129, USA
majdalany@lm-mgmt.com

Abstract. Established in 1988, the Transaction Processing Performance Council (TPC) has had a significant impact on the computing industry's use of industry-standard benchmarks. These benchmarks are widely adapted by systems and software vendors to illustrate performance competitiveness for their existing products, and to improve and monitor the performance of their products under development. Many buyers use TPC benchmark results as points of comparison when purchasing new computing systems and evaluating new technologies.

In this paper, the authors look at the contributions of the Transaction Processing Performance Council in shaping the landscape of industry standard benchmarks – from defining the fundamentals like performance, price for performance, and energy efficiency, to creating standards for independently auditing and reporting various aspects of the systems under test.

Keywords: Industry Standard Benchmarks.

1 Introduction

Originally formed in 1988, the Transaction Processing Performance Council (TPC) [1] is a non-profit corporation focused on defining database processing benchmarks and disseminating objective, verifiable performance data to the IT industry. The TPC

R. Nambiar and M. Poess (Eds.): TPCTC 2011, LNCS 7144, pp. 1–9, 2012.

was originally founded in response to a growing trend at the time, affectionately known as "benchmarketing." Effectively, this was the not-so-uncommon practice of vendors to publish amazing claims based on their own performance data in order to increase sales. Without independent and objective oversight, a number of vendors created highly suspect workloads and test environments while often ignoring crucial operational and sometimes even "correctness" requirements in order to improve the market's perception of their product.

"Benchmarketing" effectively enabled vendors to exaggerate performance and even reliability claims in order to boost sales. The need for a vendor-neutral standards organization that focused on creating and administering fair and comprehensive benchmark specifications to objectively evaluate database systems under demanding, but consistent and comparable workloads, quickly became apparent. Several influential database academics and industry leaders began working to establish an organization charged with leading the effort to impose order and consistency to the process of benchmarking products fairly and objectively – this effort ultimately culminated in the formation of the TPC.

Over the years, both vendors and end-users have come to rely on TPC benchmarks to provide accurate and dependable performance data that is backed by a stringent and independent review process. Vendors publish TPC benchmarks to illustrate performance competitiveness for their products. In addition, many vendors use TPC workloads internally to improve and monitor release-to-release progress of their products using TPC-C and TPC-H benchmarks. End-users use TPC benchmark results as a reliable point-of-comparison when considering new technologies and purchasing new computing systems.

The key to providing end-users with the promise of reliable and comparable results across both hardware and database systems starts with a well-defined specification to ensure consistency in workload and measurement. Although some might argue these specifications are too large and detailed, it is precisely this which prevents vendors from "bending" the rules to their advantage. To ensure this, TPC benchmark publications mandate extensive documentation of the configuration and benchmark process which are carefully vetted and certified by a TPC-certified and independent Auditor before it can be released as a formally approved TPC benchmark result.

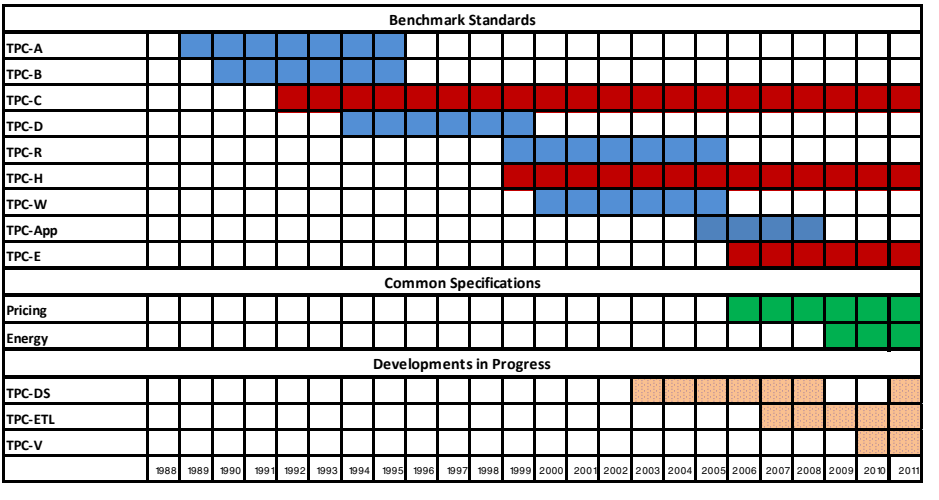
A key innovation the TPC popularized was the notion of Price/Performance. While vendors could often reach bigger and bigger performance results simply by adding more capacity or faster components, price/performance acts as a counter-balance to provide transparency to the cost of getting that level of performance. To enforce consistency in the costing aspects of these tested solutions, the TPC developed a Pricing Specification [2] designed to ensure uniformity between benchmark results. Auditors must also validate that each benchmark follows the requirements set forth in the pricing spec to ensure this. The pricing specification sets guidelines for how vendors must price the hardware, what hardware components must be included, the rules for licensing of all the software used in the benchmark, and the contract costs for three years of maintenance and support for all hardware and software.

In recent years, energy efficiency has become one of the leading factors in evaluating computer systems. To address this shift, the TPC has developed the Energy Specification [3], intended to help buyers identify the energy efficiency of computing systems in addition to the performance and price/performance requirements. Like the TPC Pricing Specification, the TPC Energy Specification is a common specification ensuring consistency across all TPC benchmark standards currently in use, including TPC-C, TPC-E and TPC-H specifications.

To better understand the TPC's contributions to the industry, let's explore the different benchmark specifications.

2 TPC Benchmark Standards

Over the years, TPC benchmarks have raised the bar for what the computing industry has come to expect in terms of benchmarks themselves. Though the original focus has been on online transaction processing (OLTP) benchmarks, to date the TPC has approved a total of nine independent benchmarks. Of these benchmarks, TPC-C, TPC-H and TPC-E are currently active, and are widely being used by the industry. TPC-ETL, TPC-V and TPC-DS are under development. The timelines are shown in Figure 1.



Current industry standard benchmarks include TPC-C, TPC-H and TPC-H, each of which addresses distinct industry requirements. TPC-C and TPC-E are standards for benchmarking transaction processing systems, while TPC-H is the standard for benchmarking decision support systems. The longevity of these benchmarks means that hundreds of results are publicly available over a wide variety of hardware and software platforms.

The top contribution of the TPC is defining the fundamental metrics that the industry uses to analyze and compare computer server technologies. All TPC results

must report three primary metrics; Performance, Price/Performance and Availability Date. All three primary metrics must be reported for a valid TPC result.

Performance is a measure of the throughput of database transactions being performed on a System Under Test (SUT) in a given operating environment. Each benchmark standard has a defined performance metric; TPC-C uses tpmC (transactions per minute), TPC-H uses QphH (queries per hour) and TPC-E uses tpsE (transactions per second). TPC performance results are widely used in the industry to analyze server performance and compare the performance of various vendor offerings. The performance must be reported in, or derived from, measured values. The result has to be repeatable. The use of estimated values in describing a TPC result is strictly prohibited.

Price/Performance is a measure of the cost of delivering the stated performance. The motivation for price/performance was driven by the need for sponsors to use configurations that are commercially viable. This metric has been used widely for purchasing decisions, especially in a highly competitive market place, where the most effective use of resources is a key objective. The Pricing Specification gives guidelines on how the Total Cost of Ownership (over three years) of the SUT is calculated. The TCO is composed of the line item costs of the hardware and software components, based on the SKUs that ship with the systems plus the cost of a three year maintenance service. The TPC pricing specification defines the process that enables these costs to be verified for accuracy.

Availability Date, the third TPC primary metric, defines the vendor's commitment in delivering the product. Having the option of using a future availability date enables vendors to preannounce, and hence generate demands for, their products. The Availability Date as per TPC definition is when the all the components of the SUT are orderable and are being shipped to customers. The Availability Date must be within 185 days of the date when the result is submitted to the TPC.

Introduced in 2009, the TPC-Energy specification defines the methodology and requirements for measuring and reporting energy metrics. TPC-Energy metrics are optional and are not required to publish a TPC benchmark result. Watts per Performance was a metric that was inspired by the realization that the high performance of a SUT usually comes at a cost of high power consumption. With the ever rising energy costs for data centers, a measure of the power a configuration consumes is relevant business information for IT managers. A key objective of this metric is to encourage and spur the development of power efficient computer technologies.

TPC benchmarks have permanently raised the bar; vendors and end users rely on TPC benchmarks to provide real-world data that is backed by a stringent and independent review process. The main user and vendor benefits of TPC benchmarks are listed below:

- Cross-platform performance comparisons. TPC benchmarks enable server configurations and solution offerings to be compared. The ability to verify vendor marketing claims is a key contribution by the TPC to the industry. By

providing a basis on which server platforms are compared, the TPC has encouraged and driven better performing technologies.

- An objective means of comparing cost of ownership. The TPC has been the most successful benchmarking group in developing a standard means of comparing the price and price/performance of different systems. All TPC testing requires vendors to detail their hardware and software components, along with the associated costs and three years of maintenance fees, in order to provide the industry's most accurate price and price/performance metrics.
- An objective means of comparing energy efficiency. The TPC Energy metric provides an additional dimension to computing systems' performance and price. As with the TPC's price/performance metrics, which rank computing systems according to their cost-per-performance, the TPC Energy metric ranks systems according to their energy-consumption-per-performance rates.
- Complete system evaluation vs. subsystem or processor evaluation. The TPC benchmarking model has been the most successful in modeling and benchmarking a complete end-to-end business computing environment. This has helped TPC benchmarks gain recognition as credible, realistic workloads. Most past and many current benchmarks only measure the hardware performance (including processor and memory subsystem). TPC benchmarks have led the way in developing a benchmark model that most fully incorporates robust software testing.
- Peer review and challenge. TPC results are used widely within the industry and the TPC has defined processes that ensure that these results are credible and compliant with the benchmark specification under which they are published. All results are checked by an independent TPC-authorized Auditor for accuracy and compliance with the benchmark specification before they can be published. A result's sponsor must publish an Executive Summary (ES) and Full Disclosure Report (FDR) detailing how the SUT was measured. Both these documents are available to the public. TPC member companies can review these documents and raise a challenge if they find any inconsistencies with the prevailing TPC policies or specifications. The Technical Advisory Board (TAB) considers these challenges in a timely manner and recommends a course of action to the TPC General Council. If a result is found non-compliant, it is withdrawn. The use of TPC results to make false and unverifiable marketing claims (benchmarketing) is strictly prohibited and can lead to a TPC Fair Use violation. This can subsequently lead to a reprimand and/or fine.

3 Defining a Level Playing Field

TPC benchmarks provide a credible way to comparatively evaluate the price, performance and energy requirements of complete systems, subsystems and/or processors. To make this possible, the TPC has taken considerable efforts to establish

a level playing field, in which end-users and vendors can agree on an objective means of comparing disparate computing architectures.

Two key components to the TPC's success include the organization's rigorous benchmark result auditing process and the pricing component of existing TPC benchmarks. These items are described in detail below.

3.1 Auditing Process

The TPC's stringent auditing process has been integral to the organization's success as a leading publisher of industry-standard benchmarks. Independent TPC-certified Auditors verify all benchmark results as a prerequisite for publication. Organizations performing benchmark tests are required to fully document the system components, applications under test and benchmark procedures. This full disclosure makes it possible to question and challenge each result, and ensures that all published results are both credible and verifiable.

Even after a benchmark result has been published, the TPC encourages a 60-day Peer Review process. During the Peer Review, every member organization in the TPC has the right to challenge the published result.

A comparison of the TPC's auditing process to that of other industry-standards organizations is provided at the end of this section. First, however, the TPC's auditing process is outlined below in further detail for added clarity:

- Verifying the compliance of all components in a SUT, including software programs, hardware configurations, purchase and maintenance pricing, etc.
- Ensuring that the methodology used to implement the benchmark tests produces results that demonstrate compliance.
- Verifying the compliance of benchmark execution by examining the results produced.
- Encouraging comment: The establishment of an audit protocol allows Test Sponsors and Auditors to document, in detail, a required set of steps which produces the specified benchmark results. The protocol also documents test methodology and the resulting test data, which is captured and communicated to the Auditor.
- Verifying the compliance of the result, based on applicable Technical Advisory Board (TAB) and General Council (GC) rulings. Additions to the audit process may be required if there are outstanding issues that have not been previously covered.
- The Test Sponsor is responsible for attesting to the veracity of all information disclosed to the TPC Auditor and in the Full Disclosure Report (FDR).
- The Auditor may choose to examine and test disclosed information at his/her discretion. The Auditor's focus is on verifying the methodology used for reaching compliance, rather than verifying the information disclosed by the Test Sponsor.

The TPC's auditing process differs from other organizations involved in creating and publishing benchmark results. The Standard Performance Evaluation Corporation (SPEC), for example, emphasizes a Peer Review process after publication in lieu of auditing benchmark results independently. This is intended to help improve consistency in the understanding, application, and interpretation of SPEC benchmark run rules. Critically, although SPEC reviews results and accepts them for publication on its Web site, the results themselves remain the responsibility of the tester. This stands in contrast to the TPC, which makes substantial efforts to ensure benchmark results are independently certified prior to publication.

Like the TPC, the Storage Performance Council (SPC) utilizes both Peer Review and independent auditing. An SPC benchmark measurement becomes a new benchmark result upon successful submission and completion of the SPC audit process, which is required. The submitted result is then given the status "Submitted for Review" for a minimum of 60 days, during which time the Peer Review occurs. Like the TPC's Peer Review process, the SPC Peer Review allows members an opportunity to review the details of the benchmark result and raise any compliance issues. If there are no issues raised during this period, the status of the benchmark result changes to "Accepted." If, however, the SPC result is found to be non-compliant during the Peer Review, the benchmark result must either be withdrawn or revised prior to additional review.

3.2 Pricing

The TPC-Pricing specification is designed to guide both customers and vendors implementing TPC benchmarks. Additionally, the specification directs TPC Auditors on what is acceptable pricing for the purposes of publication. The pricing methodology reflects the purchase price of the benchmark SUT, software licensing used in the benchmark and the contracts for maintenance.

The TPC-Pricing specification also establishes an availability metric, which provides information on whether a specific benchmark configuration can be purchased immediately or if some of the components of the configuration are not immediately available. The availability requirement limits the length of time before a promised result must be fully available. Ideally, all systems would be available immediately upon publication, but the TPC must balance the benefits of allowing sponsors flexibility in showcasing systems where one component may not be available, and currently allows 185 days from the date of publication – although most results are available immediately or within a few weeks.

To meet the requirements of being fair, honest and comparable, while allowing for a variety of pricing and business strategies, the following requirements exist for the pricing information across all TPC benchmark publications:

- Pricing must be based upon a pricing model that the sponsoring company employs with existing customers.
- The published price must be a price that any customer would pay for the priced configuration. In a competitive environment, aggressive discounting

may occur in certain situations, such as sales or closeouts. Since these situations are unique, they do not meet the requirements of the TPC-Pricing specification. Therefore, the pricing model employed for TPC benchmark publications may not represent the best or lowest price a customer would pay.

- The methodology used must generate a similar price for a similar configuration for any customer. The pricing model must represent the pricing that could be obtained by any customer in a request for bid to a single vendor. Situations that occur when requests for bids go out to multiple vendors, and then those bids are used in negotiations to get a better price, are not represented.

Benchmark sponsors are permitted several possible pricing models to construct a price for their configuration. The pricing models used must adhere to TPC disclosure requirements. Competitors often try to confirm price accuracy by calling into sales offices anonymously and attempting to purchase an actual system.

4 A Look Ahead

The information technology landscape is evolving at a rapid pace, challenging industry experts and researchers to develop innovative techniques for evaluation, measurement and characterization of complex systems. The TPC remains committed to developing new benchmark standards to keep pace, and one vehicle for achieving this objective is the sponsorship of the Technology Conference on Performance Evaluation and Benchmarking (TPCTC). With this conference, the TPC encourages researchers and industry experts to present and debate novel ideas and methodologies in performance evaluation and benchmarking.

The first TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC2009) [4] was held in conjunction with the 35th International Conference on Very Large Data Bases (VLDB2009) in Lyon, France during August 24–28, 2009, supported by the TPC in a silver sponsor role. The paper acceptance ratio was 47%. The conference was keynoted by Mike Stonebraker, recognized as one of the top five software developers of the 20th century and an adjunct professor at the Massachusetts Institute of Technology. The formation of TPC's Virtualization working group (TPC-V) was a direct result of the papers presented at this conference. Proposals like dependability aspects are under consideration for future benchmark enhancements. The conference proceedings have been published by Springer-Verlag, and are available via the following URL: http://www.springer.com/computer/hardware/book/978-3-642-10423-7.

The second TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC2010) [5] was held in conjunction with the 36th International Conference on Very Large Data Bases (VLDB2010) in Singapore during September 13-17, supported by the TPC in a silver sponsor role. The paper acceptance ratio was 58%. The conference was keynoted by C. Mohan, IBM Fellow at IBM Almaden Research Center in San Jose, who is recognized worldwide as a leading innovator in transaction management. There are several new benchmark ideas, enhancements to existing benchmarks and lessons learnt in practice presented at this conference. The

conference proceedings have been published by Springer-Verlag, and are available via the following URL: http://www.springer.com/computer/communication+networks/book/978-3-642-18205-1.

With the third TPC Technology Conference on Performance Evaluation and Benchmarking (TPCTC2011) proposal, the TPC encourages researchers and industry experts to submit novel ideas and methodologies in performance evaluation, measurement, and characterization. Authors are invited to submit original, unpublished papers that are not currently under review for any other conference or journal. We also encourage the submission of extended abstracts, position statement papers and lessons learned in practice. The accepted papers will be published in the workshop proceedings, and selected papers will be considered for future TPC benchmark developments.

Areas of Interest:

- Appliance
- Business Intelligence
- Cloud computing
- Complex event processing
- Database performance optimizations
- Green computing
- Data compression
- Disaster tolerance and recovery
- Energy and space efficiency
- Hardware innovations
- High speed data generation
- Hybrid workloads or operational data warehousing
- Unstructured data management
- Software management and maintenance
- Virtualization
- Very large memory systems
- Lessons learnt in practice using TPC workloads
- Enhancements to TPC workloads

Acknowledgements. The authors would like to thank the past and present members of the TPC for their contribution to specifications and documents referenced in this paper.

References

1. Transaction Performance Council website (TPC), `http://www.tpc.org`
2. TPC Pricing Specification, `http://www.tpc.org/pricing/default.asp`
3. TPC Energy Specification, `http://www.tpc.org/tpc_energy/`
4. Nambiar, R., Poess, M. (eds.): TPCTC 2010. LNCS, vol. 6417. Springer, Heidelberg (2011) ISBN 978-3-642-18205-1
5. Nambiar, R., Poess, M. (eds.): TPCTC 2009. LNCS, vol. 5895. Springer, Heidelberg (2009) ISBN 978-3-642-10423-7