Price and the TPC

Karl Huppler

IBM MS XQK
3605 Highway 52 North
Rochester, MN 55901 USA
huppler@us.ibm.com

Abstract. The value of a benchmark metric is directly related to how relevant it is to the consumer.

The inclusion of a price/performance metric and an availability date metric in TPC benchmarks has provided added value to consumers since the first TPC benchmark was introduced in 1989. However, over time the relative value of these metrics has diminished – both because the base price of hardware and software comprises a smaller fraction of the total cost of ownership (TCO) and because TPC pricing requirements have not kept pace with changes in the industry. This paper aims to

- Highlight the strengths provided by including price/performance and availability metrics in benchmarks.
- Identify areas where the relative value of these metrics has diminished, over time.
- Propose enhancements that could return them to provide high value to the consumer.

Many of the ideas in this paper are a result of nearly a decade of discussions with many benchmark experts. It would be difficult to identify an originator for each specific suggestion. However, it is nearly certain that this is the first comprehensive list where this collection of ideas is presented.

Keywords: Price/Performance, TCO, Total Cost of Ownership, Benchmark, TPC.

1 Introduction

When the Transaction Processing Performance Council brought out its first benchmark, TPC Benchmark A (TPC-A), in November of 1989 it introduced a novel approach that has provided value for over two decades. Not only was a strict methodology developed for implementing and measuring a performance benchmark to achieve a qualified throughput capacity for a computer system, but rules for defining a comparable price were developed and the values of price/performance and the date of availability of the total system were raised to be equal metrics along side the performance capacity metric.

These requirements provided substantial value to the consumer – by ensuring that the configurations that were measured were real configurations for which orders could

R. Nambiar and M. Poess (Eds.): TPCTC 2010, LNCS 6417, pp. 73–84, 2011.

be placed and deliveries received, and by discouraging test sponsors from overloading their system with expensive components in order to achieve the highest performance score.

For availability, there were two basic rules:

1. All hardware and software had to be orderable on the day the benchmark result was announced (This generally meant that all hardware and software was also already announced to the general public.)
2. All hardware and software had to be deliverable within six months of the day the benchmark result was announced.

For price, there were three basic rules:

1. The price had to include all hardware and software for the entire configuration, including system, storage, end-user work stations, switches, software to execute the benchmark and software to develop the benchmark.
2. In an attempt to reflect part of the cost of ownership, the price had to include 5 years of maintenance for both hardware and software.
3. The price had to be one that any customer could expect to pay

In the TPC's early years, steps were taken to strengthen the value of TPC benchmark results, including the price and availability metrics [1]. The TPC's Technical Advisory Board (TAB) was tasked with evaluating claims that benchmarks were not executed correctly. TPC Policy wording was generated to encourage fair marketing practices where TPC benchmark results were concerned. An audit process was instituted to require an independent attestation of the validity of results. Benchmark language was created to preclude the use of specialized hardware or software that was developed solely for the purpose of enhancing the benchmark result.

1.1 Eroding Strengths of the Original Metrics

Throughout the past two decades, four key trends have caused the effectiveness of these metrics to diminish.

1. Advances in technology have shifted the way that hardware and software providers interact with their customers, making some of the assumptions made in 1989 less relevant
2. Similarly, advances in technology have allowed the creation of very complex consumer applications, meaning that the relatively simple benchmark applications require memory and storage configurations that are much larger than what is installed in a typical consumer environment
3. New methods of computer solution delivery have been developed that are not included in the TPC's specification or policy language
4. In addressing these issues, the TPC has tended to relax requirements standards to accommodate an industry change rather than strengthening them.

1.2 Proposed Enhancements Included in This Paper

It is worth noting that, while the descriptions of the Price/Performance and Availability Date metrics and their historical changes are based on research and fact, the proposals

to enhance the impact of these metrics are editorial in nature. They are the result of collective experiences in the benchmarking environment, but they are essentially opinion and are presented without proof. The intent of this method is to initiate conversations on these topics within the TPC and invite public feedback, as appropriate.

2 Availability Date

We begin with this metric because it is less complex than the price/performance metric and we feel that the proposals for enhancing the metric are fairly straightforward.

As noted in the introduction, it is to the TPC's credit that this is a "metric". Benchmark results are used in marketing materials and in consumer purchase decisions. It is important for the consumer to understand whether the promised performance is available right away, in 2 months or in 6 months. The TPC requires this metric to be displayed whenever TPC benchmark results are used in public materials.

Initially, the availability requirement for TPC benchmarks was that all hardware and software must be orderable from the day the benchmark result is announced and must be available for delivery within 6 months. The declared Availability Date Metric was that date when all components of the benchmark configuration could be delivered to a customer.

Three things have changed within the industry and/or within the TPC:

First, the delivery process for some software changed. Companies no longer held orders on the books for long periods of time. Instead, they chose to only allow customers to order the software when it became ready to be shipped – often via a web download. The TPC's response was to say "OK, hardware has to be orderable from day one, but for software, you only have to describe how it can be ordered by the availability date." This response was natural, but had the effect of relaxing the requirement.

This created a disparity between the way hardware availability and software availability were treated, so the second change was to say: "OK, we'll treat everyone equally, but if your product is not orderable, you have to at least say how it can be ordered at a later date and show that the same process could be used to order a similar existing product, today." This response was fair and reasonable, and the TPC deserves credit for adding the requirement that an established order process must exist, even if a component isn't orderable until a later date. However, the net affect was to relax the prior requirement.

Finally, new development processes for both hardware and software have caused the rate of product turnover and new product delivery to accelerate. Very few consumer decisions are made based on what is available in 6 months – many are made based on what is available "now". This shift has not yet been addressed by the TPC.

2.1 Availability Date – Proposed Enhancements for the TPC

An immediate approach to strengthening the Availability Date metric lies in the hands of consumers, analysts and trade press. The approach is simply to give attention to this metric; to highlight when the products for a benchmark result are already available or will soon be available and contrast that with products that won't be available

for 5-6 months. If consumers insisted on only using benchmarks that had fully available configurations, more benchmark results would be published on available product sets.

However, the main tenant of this paper is to focus on what the TPC can do to enhance its metrics. The strength of the availability metric can be improved with two simple changes. The challenge with both is that they will remove some freedom that benchmark sponsors have enjoyed.

1. Change the availability date window to 90 days. The benchmarks of the Standard Performance Evaluation Corporation (SPEC) typically require benchmarked products to be available within 90 days, demonstrating that a 90-day rule works, even though it is "nice" to have the 185-day buffer (extended from 6 months to ensure that it would be at least six months, regardless of the start date).
2. Require that all products used in producing a result be orderable on the day the result is announced – period. The TPC's software maintenance requirements already include support for product updates, so if the level of software needed to support a result is not immediately available, all that should be needed is to be able to order the predecessor level. If it is a new version that is not yet orderable, then the upgrade price from the prior version to the new version should be included. If it is a brand new product that has never been orderable, before, then it shouldn't be included in a benchmark result until the provider has enough confidence in it to allow it to be ordered. A corollary to this is that all products used in producing a benchmark result should be described to consumers somewhere other than in the benchmark result disclosure.

3 Hardware Pricing

From a surface view, the task of generating a price for a hardware configuration appears to be very straightforward: There is a list of physical components that are required to measure the benchmark; There is a price associated with each component; Multiply the price and the quantity of each component and add them together - - Trivial if all customers expect to pay "list price", or even if all suppliers expect to offer similar discounts. However, this is hardly the case. Some suppliers list their products with fairly high profit margins and frequently offer deep discounts to their customers. Some suppliers discount very little, but price their offerings very competitively. Some suppliers tend to sell their product through resellers, where the "list price" of the supplier may be 20% higher than the list price of the reseller.

To deal with this, the TPC has a good set of rules to ensure that prices are meaningful and comparable. Many of these also apply to software and maintenance, but are listed here, first:

- The entire system under test, including storage and all servers, must be priced
- Discounts are allowed, but the basis of the discount must be described
- Both prices and discounts must be available to any customer

- Comparisons of prices must use the complete price, not component subsets, because there is no guarantee that the components are used in the same way in two different benchmark results

However, over time changes have been made to the specifications and changes have occurred in the industry that reduce the overall importance of this portion of the price. In particular:

- Where TPC-A and TPC-C once required pricing of workstations as part of the "total" system, the number of simulated users grew to be so large that the workstation price overwhelmed the price of the rest of the system, so it was eliminated from the requirement. This made the actual price more relevant, but it relaxed the "total system" requirement and caused the price per unit of work to be a smaller physical value.
- TPC-C once required pricing sufficient storage to contain 180 days of archived data, but the benchmark transactions were so efficient that this required far more storage than a typical configuration would need, so the amount of space was reduced to 60 days of archived data. Again, this served to make the price more relevant, but caused the price/performance metric to have a smaller physical value.
- Benchmark applications have become highly optimized in comparison with typical consumer applications. As a result, configurations needed to support the benchmarks, even with the adjustments listed above, have far more disk and memory than typical "customer" configurations. Discounts that can be offered on such configurations could potentially be deeper than some customers would pay for more typical configurations.
- Recognizing that a specific sales organization could deliver different pricing than a more general pricing method, such as a web tool, all that is required is to identify a pricing source that any customer has access to (1-800-BUY-ACPU, for example). This method has been casually called "TPC Price Desk" The current availability requirements dictate that such a method be allowed, since benchmarks are allowed to be published prior when configurations can be ordered through a normal sales channel.

The first two of these are really reflective of the third point. TPC benchmarks are focused on a specific set of database operations. They are touted as "full system", but they do not include the complete path of a "full application". In fact, while the database activity is fairly robust, the actual "application" layer is as slim as the benchmark implementers can make it – in order to achieve more database performance. In reality, processor configurations used to measure these "database subsystem" benchmarks would process a fraction of the data processed by the benchmark, because there is a great deal of other work that they must accomplish outside of the database subsystem. Consequently, the amount of memory and storage configured to support the database is about an order of magnitude more than is typically included in a price quotation for a consumer configuration.

3.1 Hardware Pricing – Proposed Enhancements for the TPC

If we accept that the benchmark configurations are exaggerated, and that this allows for pricing that, while be normal for these configurations, could be abnormal for more typical configurations, it follows that to the TPC should find a way to price more conservative configurations. There are three ways to accomplish this:

1. Patch the benchmarks to reduce requirements, as was done with the first two bullets, above. This provides a temporary improvement, but it is cosmetic, at best.
2. Change the benchmark to become more robust and require more processing resources relative to the other resources in the configuration. This is an admirable approach, and one that partially succeeded with the introduction of TPC-E, but benchmark development takes a long time and this would diminish the data base focus that is a mainstay of TPC benchmarks.
3. Define a subset configuration to be priced, based on the measured configuration. Each benchmark would have its own rules. For example, TPC-C might require pricing a database server with 1/8 the memory used in the benchmark and 1/8 the total number of disks.

This last solution is radical, in that it diverts from the standard rule of "price what you measure, measure what you price." However, it resolves the challenges listed here quite nicely: Configurations would be of a size that customers might actually consider buying. Published benchmark prices would be self-verifying, because customers would go to their price source of choice and demand a similar or better price.. Furthermore, the methodology associated with this can be adjusted in a new version of the benchmark to reflect current buying practices and technology.

The fundamental point is to price configurations that have meaning to the consumer, thereby increasing the relevance of the price/performance metric.

4 Software Pricing

Software presents a particular challenge in pricing requirements, in that customers do not buy software. They pay a license fee for the privilege to use the software. There are often a number of options available to consumers:

- license a specific number of users to exercise the software, where a small number of sessions are expected to use the software
- license a specific number of processors or processor-cores on which the software will be exercised, where the use of the software is unlimited, as long as it is run on no more physical resources than licensed
- purchase a perpetual license, where the licensee is allowed to use the software "forever", even after maintenance support for the software is dropped
- purchase a term-limited license, where the licensee is not allowed to use the software once the term expires, unless the license is extended
- purchase of a utility-based license, where charges are based on the frequency with which the code is exercised.

Any of these options may be viable for a particular consumer scenario. However, a performance benchmark introduces the challenge of producing comparable results.

Utility-based licenses would be difficult to enforce in a benchmark environment. They are typically oriented to a very large number of potential users, each only accessing the computer for a brief period. This is in contrast to benchmark environments when the computer is running at maximum capacity.

Examining user-based licensing, the benchmark application tends to be simpler and more stream-lined than a typical consumer application, the number of active sessions needed to drive a system to full capacity may be far fewer than the number that may access similar software in a consumer environment. This is compensated in TPC-E, where the number of users licensed must be equal to the tpsE throughput value and in TPC-H, where the number of licensed users must be 10 times the number of query streams executed – with the result that almost all benchmark results use licenses that are priced by number of processors or cores.

Price by cores also has challenges. Seldom will a large system be used exclusively for database activity, and seldom will it be executing at 100% of system capacity, Consequently, there are more processing cores active during the benchmark than in almost any consumer scenario. Furthermore, all processor cores are not created equally, so many software vendors establish rules for each of several architectures that either charge less per core for some architectures than others, or that create a smaller number of "chargeable cores" than what is in the configuration. However, this is also how the licenses are presented to actual consumers, so apart from the potential for a company to provide lower prices for software running on "preferred" hardware than on other hardware, it may be as close to reality as possible.

Finally, there is the difference between perpetual and term-limited licensing. Term-limits are more similar to a lease of a license than a purchase. Note that, in the hardware discussion, the term "lease" did not surface - - This is because the TPC has successfully barred hardware leases from being used in benchmark pricing. This does not mean that a hardware lease is not a viable option for a consumer – only that it was deemed not to be comparable with hardware purchases. Even after a hardware product has been fully depreciated, the consumer has the option of retaining and exercising the hardware – This is not the case with term-limited licenses.

4.1 Software Pricing – Proposed Enhancements for the TPC

The pricing of software could be made to be much more comparable with the implementation of three enhancements:

1. Treat hardware and software consitently by requiring a "perpetual" purchase of the entire configuration, retaining some residual value for as long as the consumer desires.
2. Rather than try to compensate for the simulated number of users in a configuration, the TPC should simply disallow per-user-based or usage-based licenses and require perpetual software licenses on a per-core or per-processor basis.
3. Apply a "reality adjustment" that is similar to the proposal for storage and memory in the hardware examples. Each benchmark committee should determine the ratio of number of cores to be licensed relative to the number measured, perhaps reducing this be a factor of 2-3.

As with hardware, the latter suggestion strays from the concept of "Price what you measure; measure what you price." It is proposed in the spirit of delivering priced configurations that are consumer-based, rather than benchmark-based, to encourage pricing of configurations that are relevant to consumers.

5 Maintenance Pricing

Maintenance of hardware and software is a critical part of data center operations, and a significant component in the total cost of ownership for a computing solution. Maintenance, or the broader category of "service," is an area that is uniquely customizable to fit the consumer's needs. Almost all customers require warranty-level maintenance on their products to fix items that fail or have defects, but most customers also require a level of additional support that helps them to make optimal use of their investment. This includes such things as software upgrades, consultation, education offerings, predictive analysis and myriad other options. Because customers' needs differ, product suppliers tend to offer a variety of options.

When the TPC first defined rules for pricing, many systems were used for well over five years before being replaced with newer technology. Five years was also the fastest that the United States Internal Revenue Service and United States Generally Accepted Accounting Principles (GAAP) allowed for depreciation of assets. Similar rules were established with the International Financial Reporting Standards (IFRS) and the International Accounting Standards (IAS) This prompted the TPC to require maintenance support for a five-year period from the initial publish date. Over time, it appeared that the rapid turn-over of technology would prompt depreciation schedules to be shortened to three years, and the TPC adjusted to use that schedule.

Hardware maintenance requirements for TPC benchmarks are fairly straightforward and, in the opinion of this author, fairly close to consumer requirements. In summary, there are two methods of maintenance allowed for hardware by the benchmark specifications:

1. 7X24 support with 4-hour response time to begin working a problem, continuing work until the problem is resolved
2. For customer-replaceable parts, the option to price an on-location supply of 10% extra parts, with a mail-in replacement offering for failed parts.

Although other hardware maintenance offerings are provided by many vendors, these requirements are representative of what a consumer might choose in an environment that supports critical run-your-business applications. Not all vendors offer the second option, but virtually all vendors that sell into mission-critical environments offer the first of these.

It is more difficult to define specific levels of software support, because software also has "soft" problems – namely user errors, software configuration questions, implementation and optimization concerns, and the like. Furthermore, many software providers handle these support items in a different way. To provide a set of requirements that are common among most software vendors, the TPC chose a "lowest common denominator" approach, requiring only the most basic maintenance for software. The software maintenance requirements can be summarized as:

1. 7X24 support with a 4-hour response time to acknowledge that a software bug exists and begin working on it - - - but with no specific commitment to resolve the bug in any given amount of time
2. Access to maintenance update software, such as Service Packs, that include rolled-up groups of fixes to previously resolved software bugs.

Areas that are not included and are important to almost all customers:

- Functional software updates (new releases)
- Operational guidance
- Customer Education

The result is that often the maintenance offerings that are priced in a benchmark satisfy only the bare minimum requirements of the benchmark and do not reflect the needs of a typical consumer.

5.1 Maintenance Pricing – Proposed Enhancements for the TPC

Three enhancements would bring TPC requirements for maintenance and support to a point where it would reflect relevant information for the consumer:

1. It is first worth noting that depreciation schedules have remained at five years, for the most part. A reflection of this can be found in the United States IRS form 4562, used to document depreciation for tax purposes in the United States [2].Consequently, it is recommended that maintenance prices be extended to the TPC's original five-year requirement.
2. However, we also note that the initial purchase of the equipment is a capital investment, while the support costs are considered to be expenses – most often spread on an annual, quarterly, or monthly basis. Although the total price/performance metric should reflect the full five-year costs of initial purchase and maintenance, there should be secondary metrics that call out the initial purchase price and the on-going yearly maintenance after the first year. It should be valid to compare benchmark results based on initial purchase price, annual maintenance and total 5-year price.
3. Most important and most difficult to define: The requirements for software support should be raised to a level that is relevant to a consumer whose business depends on the products installed on the system(s). Support requirements should include the current TPC requirements, plus ongoing product updates, operational guidance and some level of customer education. Each component of maintenance should be priced for no more than one year, multiplying that value by the number of years needed (assuming some coverage is included in the purchase price for the first year) to create a five-year maintenance cost. The challenge will be drawing a line between "operational guidance" and "consultation services". Most companies offer both, but do not always use the same criteria in shifting from the moderately priced, almost always purchased option and the more expensive level of support that would include design guidance, optimization and other services that might more typically be fee-for-service options.

6 The Relationship between the "TPC Price" and TCO

The TPC does not claim that the required price sheet for their benchmarks represents the total cost of ownership of the system, although the inclusion of maintenance requirements have prompted some to call it the "cost of ownership" without the word, "total." The following table represents many of the elements of the total cost of ownership for a computer system, along with comments on whether the cost is covered or partially covered in the existing TPC requirements or in the proposed enhanced requirements.

Element	Included in TPC-Price?	Included in Recommendations?	Comment
Initial HW cost	Yes	Yes - Improved	Improved in recommendations by requiring a more "consumer-appropriate" configuration
Initial OS, Middleware, Database SW cost	Yes	Yes - Improved	Improved in recommendations by requiring perpetual licenses on a "typical" number of processor cores
Initial Application SW cost	No	No	Often a major component of the cost, but not possible to include in a generic benchmark
HW upgrades cost	No	No	Ability to upgrade without a total replacement can be a key factor in a purchase decision, but is not included, here.
OS/Middleware/Database SW upgrades cost	No	Yes	Proposed support costs include software upgrades
Application upgrades cost	No	No	Another missing element that cannot be in a generic benchmark
HW maintenance	Yes	Yes - Improved	Improved in recommendation by extending for 5 years and providing differentiation between capital and expense costs
OS/Middleware/ Database maintenance	Yes	Yes – Much Improved	Improved in same ways as hardware, plus requiring a more robust level of support that is appropriate for most consumers with mission-critical computing environments.
Application maintenance	No	No	See other "application" entries
Application set-up, customization	No	No	See other "application" entries
DB administration	No	Minor	By including an operational level of support, a small amount of DB administration costs are also included
Systems operations	No	Minor	By including an operational level of support, a small amount of Systems Operations costs are also included
Electricity	No	No, but	This growing area of cost of ownership is covered by the TPC's new TPC-Energy metrics that can be published with each benchmark at the sponsor's option. Unlike price,

			which includes variations based on channels, discounts and other marketing-related areas, the energy component of the total cost of ownership can be quantified with verifiable measurement. [3]
Floor-space, building costs	No	No	This is quantifiable area that could be a candidate for inclusion in a future price metric.
Training	No	Yes	The recommendation for support costs includes that some level of consumer education be required in the ongoing costs.
Unscheduled down-time costs	No	No	While it is difficult to quantify, this is not only a major part of the TCO, but can be extremely disruptive to the business. A proposal in the TPC-TC '09 recommended that the TPC undertake the creation of a "dependability" benchmark. [4]
Scheduled down-time costs	No	No	If a resiliency benchmark is created, it should cover both planned and unplanned down-time

7 Summary

A benchmark metric is only as valuable as it is relevant. The price/performance and availability metrics from TPC benchmarks have been quite relevant in the past, but changes in technology, changes in other elements in total cost of ownership and changes in the benchmarks themselves have reduced the worth of these metrics.

By implementing the suggestions raised in this paper, the TPC could dramatically improve the relevance of the price/performance and availability metrics of TPC benchmarks. These suggestions should be viewed as an entire set, rather than specific individual decisions. Some of these changes would affect all vendors. Others would affect only some. With the entire collection, the relative impact to each vendor is mitigated by both the affect on other vendors and the closer approximation of consumer reality.

The proposed changes focus on

- Physical configurations that are more relevant to consumers
- Availability requirements that are closer to consumer buying practices
- Software licensing requirements that represent a typical mission-critical environment, where longevity of an application is expected
- Maintenance requirements that represent the typical levels of support that a consumer would require for a run-your-business application

These changes do not approach the total cost of ownership for a computer system, but they enhance the portions of TCO that are covered by the TPC Price Specification to more realistically represent the actual contribution to TCO for these components.

Implementing these changes will take courage and clearly generate benchmark results that are not comparable with those published today. Not implementing these

changes destines the price/performance metric, and to a lesser degree the availability metric, to reside in a realm where marketing claims are made, but the actual results are ignored.

7.1 Another Alternative

Without changes such as those suggested here, the TPC could be better off without official price/performance metrics at all. It may be that the best solution is to require that a well-defined bill-of-materials be listed, and that the list price for any products that are not publicly listed through another source be included, but to leave the computation and comparisons of price to the marketing teams, rather than the performance engineers. The advantages of this are that the comparison method would allow "for purchase" comparisons, "for lease" comparisons, "utility/cloud" comparisons, "purchase only" comparisons, "total cost of ownership" comparisons, and so on. The disadvantages of this are that the methods for comparison would be unregulated and could easily create confusing or misleading information.

References

1. Shanley, K.: History and Overview of the TPC (February 1998), http://www.tpc.org/information/about/history.asp
2. United States Internal Revenue Service, Instructions for Form 4562 (2009), http://www.irs.gov/pub/irs-pdf/i4562.pdf
3. Transaction Processing Performance Council, TPC-Energy Specification (2010), http://www.tpc.org/tpc_energy/default.asp2009
4. Vielra, M., Madiera, H.: From Performance to Dependability Benchmarking: A Mandatory Path. Springer, Heidelberg, ISDN 0302-9743 (Print) 1611-3349 (Online)