Frequently Asked Questions (FAQ)

August 2, 1999 version .99

What is TPC-W?

Introduction

This is a list of Frequently Asked Questions concerning the TPC-W benchmark. It is not intended to answer every question about benchmarking in general, or even TPC-W in particular, but it should provide a good starting point in any effort to become more familiar with TPC-W. The questions are grouped into general topics which loosely follow the structure of the TPC-W specification. Please refer to the first section below General and Administrative Questions to be sure that this document is current with the revision of the benchmark specification that you are interested in.

This document uses several terms and phrases which are specific to TPC benchmarks. It is assumed that the reader is familiar with basic TPC benchmark terminology, and has access to the TPC-W benchmark specification.

DISCLAIMER: This document is accurate to the best of our knowledge. Any discrepancy between this document and the specification represents an error in this document. The specification is the only definitive source of information, and supersedes any contradictory information found in this document.

0. General and Administrative Questions

0.1 What is the current revision of the Specification?

The current revision of the TPC-W Benchmark Specification is draft version D 5.0, released to the Subcommittee on 12 July 1999.

0.2 Who maintains the specification/FAQ?

The specification/FAQ is maintained (and periodically revised) by the Transaction Processing Performance Council (TPC). Please refer to the TPC's general FAQ for more information on the TPC. A good place to start is the TPC's home page (http://www.tpc.org). The TPC may also be reached at:

TPC
777 No. First Street
Suite 600
San Jose, CA 95112-6311
USA

Phone: (408) 295-8894
FAX: (408) 295-9768
email: shanley@tpc.org

0.3 How can I get a copy of the specification?

The latest version of the TPC-W Benchmark Specification is available on the TPC's World Wide Web server. The URL is: http://www.tpc.org/miscellaneous/TPC_W.folder/Company_Public_Review.html

The formats available in this directory include: PDF, PostScript and Microsoft Word

If you cannot access the on-line versions of the specification, a hard copy version is available from the TPC for a small fee.

0.4 What does TPC-W Measure and why is that important?

TPC-W measures the performance and price performance of computer system hardware and software used for transactional web environments such as electronic commerce, business to business and intranet. It provides a level playing field for comparison of hardware and software available to support those environments. A rigorous specification and auditing requirements are used to ensure valid comparisons.

0.5 What does TPC-W Provide that SPECWeb or WebStone or other current metrics do not?

TPC-W includes database access of data to generate dynamic web pages, a secure user interface, external secure transaction for payment authorization, scaling rules varying both number of items in web site and number of users independently. It also provides an audited price/performance metric to give the cost effectiveness of the solution with a 3 year total cost of ownership including hardware, software and maintenance.

0.6 What type of Web sites are represented by the TPC-W workload?

There are three types of web sites represented, shopping, browsing and business to business. The primary metric is a shopping site, with searching, browsing and buying functions being exercised. The browsing site does primarily browsing and searching, while the business to business site does primarily secure purchasing functions.

0.7 Who publishes TPC benchmark results?

Benchmark sponsors publish the results. They are usually vendors of the hardware or software used in the benchmark.

0.8 What is represented in the price performance calculation? What is not included?

It includes the purchase price of the hardware, runtime software, development software and maintenance for a period of three years. The cost of developing the application is not included.

0.9 What is a benchmark special?

A benchmark special is an implementation of the benchmark which uses techniques which could only be used for the benchmark, not for a real application.

0.10 Why are commercially available products required? Why is Linux a commercially available product?

Commercially available products are required to avoid benchmark specials for elements of the system which are readily available. Linux is generally available and has maintenance support packages available.

0.11 Why is availability date a required metric?

A sponsor is allowed to publish a result with an availability of some or all of the components up to 6 months in the future. These results could be compared to results which run on components available now. This is useful information for the consumer in evaluating the two results.

0.12 How was TPC-W developed

A group of 24 companies with an interest in the electronic commerce/transactional web environment, including platform, database, web server, electronic commerce package vendors worked together to define a benchmark which represented workloads relevant for their markets.

0.13 What is the customer scenario for this benchmark?

The customer, using a web browser, starts at the home page of the store, does searches and selects items to get more detailed information on, puts items in a shopping cart for purchase, enters personal information over a secure connection, provides payment information over a secure connection and approves the purchase. The customer can select additional items, do various types of searches, update the shopping cart, leave and return to the web site.

Clause 1. Web Object and Logical Database Design

1.1 Why was a bookstore store-front chosen instead of another type of retailer like CD's, or clothing or software ?

The bookstore model was chosen only as a convenience to help formulate the design of the workload. In reality, it could be anything. The system components exercised would be the same.

1.2 Why are there only eight tables in the logical design? Certainly a retail store front logical design is not this simple.

It is true that in reality, the database structure of a retail store on the web would be much more complex. The goal of the design of the database was to support a workload that would exercise the components of a system in an electronic commerce environment. The resulting design with just the eight tables accomplishes this goal, while keeping the implementation simple.

1.3. Why are the images not required to be stored in the database?

In the real world, most images are stored in filesystems. The specification allows the images to be stored either in the database or in the filesystem.

1.4. Can I add/drop/combine/split tables or columns in tables?

The base tables have to be created as specified. Tables could be partitioned, either horizontally or vertically, as long as the details are disclosed. Additional columns and / or larger fields are permitted as long as they do not enhance performance. This is to allow the use of electronic commerce products without extensive modifications.

1.5. Can I change the column datatypes (integer vs numeric vs real etc)?

No, the datatypes for the columns cannot be changed. A field definition has been specified for each column. Test sponsors are free to implement it any way they want as long as it conforms with the specified field definition.

1.6. Is there any restriction as to which columns I can index ?

No restrictions have currently been defined.

1.7. What is vertical and horizontal partitioning?

Partitioning allows the decomposition of large table data into smaller parts. Vertical partitioning is where the data is partitioned along column boundaries and data from a subset of columns of a table are grouped and stored together; horizontal partitioning is where the data is partitioned along row boundaries satisfying a given criteria and stored together.

1.8. Can the names of the tables and columns be changed (address vs addr)?

Yes.

1.9. Can my physical database span across multiple databases, database instances, and machines?

The database can span multiple instances and machines, but it must appear to the application as a single database.

1.10. Why are there five different sizes for the fixed images? And why are they so big?

The fixed sizes and the five different sizes are again a convenience for the implementation of the benchmark. It enables us to level the playing field by ensuring that all sponsors who report a result are doing the same amount of work, i.e., transferring the same amount of data. The five different sizes are an attempt to represent different media on the web within the scope of the database design. For example, a video-stream or software download is significantly more bits than the picture of a book-cover. The distribution of the sizes was arrived at after studying the page sizes at several sites.

Clause 2. Web Interactions and Workload Profile

2.1 Why should the SUT attempt to identify if the user is a known customer?

Most electronic commerce sites keep track of the way their users access the site. The information gained from this allows the site to target its users with advertising and promotional materials that match their interests and buying habits. Our requirement that the SUT attempt to identify users in some of the interactions, in conjunction with the promotional processing defined, simulates this behavior without introducing undue complexity.

2.2 If implementation of Cart is not specified, then why is it defined?

Most electronic commerce sites support the concept of a shopping cart, which may be implemented in several ways. The shopping cart definition specified is the minimum logical design necessary for this benchmark. Implementations using electronic commerce packages with more complex shopping cart functionality are also allowed.

2.3 If the implementation of Flags is not specified, then why do the interactions depend on them?

Flags are a conceptual term used in the specification to help explain benchmark functionality. The benchmark sponsor may use other techniques to implement this functionality.

2.4 Why are there multiple definitions of interactions that end with the same result [e.g., New Product, Best Seller, ...]?

Multiple interactions have been defined even though they end with the same result because they stress the SUT differently due to the different search criteria.

2.5 Why is *Cart Empty* displayed in some interactions but not in others ?

The benchmark has been designed such that the shopping cart will never be empty for the Buy Request and Buy Confirm web interactions. However, the SUT is expected to return an appropriate error message if ever a user tries to view or check out with an empty cart.

2.6 Some of the database accesses seem contrived, is this intentional and if so, why?

The interactions were designed to ensure that representative work was taking place in the SUT. This may result in some accesses seeming contrived. This was done to maximize the value of the workload within the defined set of interactions.

2.7 How is the customer password handled?

All web interactions that involve user names and passwords are secure. All passwords are stored in the database.

2.8 What is the potential benefit in the workload of including the New Customer scenario?

In the real world, electronic commerce sites do gain new customers (at some point, every customer must have been a new customer). The SUT has to do more work to keep track of whether a user is a known customer or not and create new customer records for new ones.

2.9 Since most administrative tasks are conducted during off-peak hours, what is the potential benefit to the workload by including the Admin Functions during the measurement interval?

The Admin Functions ensure that the item table is not static and hence accesses to it cannot be cached. For example, the New Products Page will have to be generated dynamically since the data it uses can be updated by the Admin Function. We wanted to include the performance effect of performing these administrative tasks on the online system.

2.10 Why was the concept of *atomic set* of operations added and what are its requirements?

The atomic set concept was added to differentiate it from database transactions. The atomic set of operations are functions which occur across web servers and external payment authorization servers, not just within the OLTP database. The interaction atomic set of functions have similar ACID properties to database transactions.

Clause 3. Database Transaction and System Properties

3.1 What are ACID properties and why are they important?

The system properties of Atomicity, Consistency, Isolation and Durability (ACID) are those that keep data which is protected by these properties from being corrupted due to interference with other concurrent users, applications or failing system components. These are important because people want to know that their data will be there tomorrow. Jim Gray explained it best, “All of the mechanisms that make databases more durable are very important. They are certainly in any database system. If you use a word processor you have a ‘quit’ button and a ‘save’ button. The ‘quit’ button is (equivalent to) ‘abort the transaction’. The ‘save’ button (means) ‘save the transaction and make it complete’. If you are drilling holes in a piece of metal you want the hole to be in the metal and the database to know that there is a hole in the metal. Transactions are important for manufacturing, they are important for all of the financial things we do, (such as) fund transfers, they are important for taking a book out of the library”

3.2 If the required tests are not sufficient to demonstrate ACID requirements, then why are the tests required at all?

It is not practical to exhaustively test ACID requirements. The tests that are specified significantly increase the confidence that the system does have the required ACID properties.

3.3 Must the ACID tests be done for every platform the TPC-W is run on?

The tests must be done for each platform. Some of the potential ACID test failures are platform dependent.

3.4 Are special transactions required to execute the ACID tests or are one of the web interactions used?

Special versions of the web interactions are used to provide control over the transaction state of the interaction.

3.5 Are the ACID tests run during the measurement intervals or are they run independently?

The ACID tests are not run during the measurement interval. The tests are run while the full transaction set is executing, and can be very invasive, including a system crash and recovery.

3.7 What is the difference between a web interaction and a database transaction?

Web interaction is from browser interface to browser interface and includes zero or more database transactions.

3.7 Why was the concept of web page consistency introduced and what are its requirements?

Web page consistency was added to assure that as changes are made to the database those changes always reflect a picture or price which always match. If the price and picture for an object change they mus both change. As database transactions update the content of the database, the web pages must display a consistent reflection of these updates: the web pages must reflect either the effect of the entire update operation or none of it.

Clause 4. Scaling and Database Population

4.1 Does the database population and size model any real-world sites?

The database population and size is based on an analysis of several real-world sites but is not meant to model any one real site.

4.2 How does the size of the database in TPC-W compare to that of the TPC-C benchmark?

TPC-C databases vary in size depending on the number of warehouses configured. Since there is no comparable table in TPC-W that scales similarly, there is no way to compare the size of a TPC-C database with a TPC-W database. For a given platform, depending on the number of items and whether or not the images are stored in the database, the TPC-W database may be larger or smaller than the TPC-C database for that platform.

4.3 Why are two variables (# of ITEMS and # of emulated browsers) used to determine the scaling of the database tables?

Unlike a physical store where the size of the inventory determines the physical space allocated, which in turn determines the number of customers it can cater to, the store on the web has to deal with the 2 issues independently. Thus, the 2 variables are used, where # of Items controls the size of the inventory and the # of Emulated Browsers controls the size of the supported customer population.

4.4 What is the rationale behind choosing the range of the cardinality of the ITEM table to be between 1K and 10M?

A cardinality of 1K seemed sufficiently small for a starting point for benchmarking purposes. A few different scale factors, with enough of a difference between them dictated the current range.

4.5 For smaller ITEM table cardinality selection (e.g. 1K), won't the entire ITEM and AUTHOR tables be cache resident?

It is possible that the Item and Author tables could be cached.

4.6 What is the rationale of choosing a 180-day database space requirement?

Configuring a database to hold 6 months of data seems reasonable.

4.7 Why does the 180-day space assume an eight hour workday?

Eight hours of peak load is roughly equivalent to 24 hours of fluctuating load typically seen in a 24 hour environment.

Clause 5. Performance Metrics and Response Time

5.1 Why are there three different transaction mixes?

Web site usage varies widely. The benchmark represents three common web site usages which have very different characteristics. These three usages are: 1. primarily browsing, 2. 'typical' shopping, 3. Primarily ordering or business to business.

5.2 How was the makeup of each transaction mix decided?

Statistics from production web sites were analyzed, a survey was distributed to gather input from targeted web sites and the effects of near term technology improvements were included in defining the mixes.

5.3 Most browsers cache web objects. How is this represented in the benchmark?

The interactions between the web browser and system under test represent the workload after taking into account the caching in the browsers. In other words, only the non-cached items are requested. A list of previously downloaded items are also maintained in the browser emulators to simulate the effect of returning to a previously cached page and selecting a new item from it.

5.4 Why doesn't the benchmark take into account the time taken to display web objects on the screen since this is a component of the response time that a user would normally experience?

Since the performance of the system under test is what is being measured, a fixed delay is incurred in the browser emulator to simulate the time taken to display the items and select new items. The browser and network performance is specifically not being measured.

5.5 Why does the SUT have to be capable of sustaining the rated throughput for only 8 hours when in a web environment the working day is typically much longer.

Full throughput is required for 8 hours, but the SUT must be able to maintain at least 30% of rated throughput for the other 16 hours. Most web environments, even though they are available for use 24 hours a day, only have peak throughputs for a portion of the day. The test system must also be available to operate for at least 2 weeks, without requiring shutdown for maintenance.

5.6 Can the RBE be script driven instead of having it generate all web interactions dynamically at run time?

The RBE can provide the interactions with any approach which meets the requirements. Since some of the interaction choices are dependent on data which is changing during the benchmark (such as search results), the 'scripts' would need to be sufficiently capable to support that.

5.7 Is there an upper limit for the duration of a measurement interval?

No, the minimum is 30 minutes, but there is no maximum.

5.8 Why must the 90th percentile response time be greater than or equal to the average response time?

It is not considered acceptable behavior for some interactions to have extremely long response times, even though other requirements may be met.

5.9 In a multi-node system there may be substantial differences in the times represented on the clocks of each node. What requirements does the benchmark have that the clocks on each component, including the RBE(s), are set to the same time?

It is necessary to be able to correlate the times of the data logged in each RBE, not to have the RBE's themselves set to the same times.

Clause 6. SUT, RBE and Network Definitions

6.1 The performance of only some system components is measured and priced—these components are collectively known as the System Under Test (SUT). What components comprise the SUT? Why was the SUT defined in this way?

The components of the SUT are the hardware and software of application servers, web servers, database servers, intra-SUT communications and attachments to external communications networks. The SUT was defined in this way to provide a controllable boundary around the system being measured. The external communications and browser for a given system are also often already in place and span a wide range of performance.

6.2 What are the main components of the system being *measured*?

The main components are the application servers, web servers, communication interface and database server hardware and software.

6.3 What components are not *measured* in the implementation?

The web browsers, payment authorization servers and internet delays between the system being measured and these components.

6.4 Network latency and browser interactions are not measured. Why?

Network latency and browser latency are normally the result of environmental factors that are beyond the scope of this benchmark. These include but are not limited to modem speed, physical network media, browser software level, time of day (as an influence on internet congestion), etc. The external network is outside the scope of the benchmark due to answers to 6.1.

6.5 Can packaged applications be used to run the benchmark?

Yes, they can as long as they do not rely on active applet elements. Flexibility was provided in data structure definition, interaction control and web page layout to allow packaged applications to be able to run the benchmark.

6.6 Are there any major restrictions on the SUT or other system components?

Yes, active elements or applets cannot be run on the RBE.

6.7 What is an RBE?

The Remote Browser Emulator (RBE) is the software component that drives the TPC-W workloads. It emulates Users using web browsers to request services from the System Under Test (SUT). The RBE creates and manages an Emulated Browser (EB) for each emulated User. The term RBE, in the TPC-W specification, includes the entire population of EB.

Clause 7. Pricing

7.1 Why is three year pricing used instead of the 5 year pricing used on other TPC benchmarks?

The recent trend among hardware vendors is to convert from 5-year pricing to 3-year pricing. The TPC-W subcommittee considers 3-year pricing to be more representative.

7.2 Why are browsers excluded from pricing?

Browsers for Internet access are typically not part of the Web Server (host) network configuration. They normally reside on a client machine owned by the end user or by the employer of the end user. As such, their purchase cost would ordinarily not be included in the price of the host system.

7.3 Why are 180 days of on-line storage required to be priced, especially if only two weeks of online storage is required on the test system?

180 days of on-line storage is representative of production systems, and must, therefore be priced. However, configuring test systems with storage capacity in excess of that required to execute the benchmark, is normally difficult for test sponsors. To accommodate this problem, the 180 days storage capacity is priced but need not be configured.

7.4 Why are products not required to execute the benchmark, such as administration, maintenance and application development included in the pricing?

Products for administration and maintenance are necessary in all production environments. The intent of this benchmark is to emulate robust 24X7 production systems executing electronic commerce applications.

Tools for application development, such as compilers, are priced if they are required to build applications running during the benchmark on the system under test (SUT). Sponsors using commercial Commerce Software Packages must include the price of these software components, which are ordinarily much more costly than the compilers and development tools. Requiring the pricing of commerce software packages but not of development tools would give greater unfair advantage to sponsors opting to develop their own software in lieu of using commercial packages.

7.5 Why is publication allowed on products which are not available for 6 months?

The TPC encourages vendors to consistently improve the performance of their products and tries to foster innovation. Allowing vendors to publish benchmark results with early product enhancements serves to further these goals.

7.6 Why allow discounts instead of requiring list pricing?

Discounts are common in the computer industry for many reasons. Discounts allow vendors to react to competition and changing market conditions. As long as discounts are for freely available products and not intended as special purpose pricing for benchmarking, they are beneficial.

7.7 Why use local pricing instead of a fixed location?

It is important that benchmark components (e.g. hardware, software) be priced in the currency of countries where the products are actually available. Otherwise, it would be possible for vendors to publish special prices for countries in which their products could not be ordered. This would create no risk for the vendor, since the products would not actually be ordered or purchased, and give an unfair price advantage.

7.8 Why require that the products be orderable when they may not be available for 6 months?

Products must be orderable to prevent sponsors from benchmarking with prototypes and ‘benchmark specials’ that were never intended to be legitimate product enhancements.

7.9 While the web allows an unlimited number of users, why is the number of users defined to be equal to the number of emulated browsers?

Although the Web theoretically allows an unlimited number of users, only a subset are actively accessing a particular Web site at any given time. Furthermore, there is a one to one correspondence between an instance of a browser and a real user.

7.10 Why are the RBE and PGE not included in the price, since it is required to run the benchmark?

The RBE is required to drive the benchmark, but is not really part of the system under test (SUT). The RBE emulates browsers, which are not normally components of the host system. See Section 8.2.

The measured components (SUT) of TPC-W are emulating a retail merchant on the Web. The Payment Gateway (PGE) is a component representing the credit card company and is not part of the retail merchant system. The PGE processes credit card payment authorization requests sent by the merchant and informs the merchant whether the cardholder is authorized to use his credit card for this purchase. Since the PGE is not part of the merchant system, it is not appropriate to include its price in the SUT cost.

7.11 While the cost of hardware, execution, and development software is included, the application development cost is not included. Why not?

It is difficult to develop a reasonable methodology for pricing application development cost, and there is no way to verify the reported price. For example, if software developer’s time is priced, how could the reported number of hours of development time be verified?

The current pricing specification favors sponsors who elect to develop benchmark software in lieu of using commercial software packages. Customers using benchmark results must take this into account when evaluating TPC-W.

Clause 8. Full Disclosure Report

8.1 What is a Full Disclosure Report and why are they relevant to the benchmark result?

The Full Disclosure Report provides the details that allows verification that the benchmark was properly executed and provides sufficient information to allow a third party to reproduce the result. It also provides more detailed performance information to allow customers to better understand the performance characteristics of the system.

8.2 What is the relationship of the Executive Summary to the Full Disclosure Report?

The Executive Summary is a brief summary of the key data from the Full Disclosure Report which has the most general interest.

8.3 *Why* are Full Disclosure Reports so large?

The benchmark is quite complex, with code implemented in various servers. To record the implementation, parameters of interest, details about how the benchmark was run, and metrics of interest is a lot of data.

8.4 How do I get a Full Disclosure Report?

Full Disclosures are available on the TPC web site (www.tpc.org). Hard copies may also be ordered from the TPC Administrator.

8.5 Am I free to copy a Full Disclosure Report?

Yes, provided the TPC trademark is included and the copyright is adhered to.

8.6 Where can I get electronic copies of programs documented in the Full Disclosure Report?

You may request them from the test sponsors, but they are not required to provide them.

Clause 9. Audit

9.1 Why does the TPC use Auditors for benchmark results?

It provides a qualified third party verification of the results. This increases the confidence that the benchmark meets the specified requirements.

9.2 Who are the auditors for TPC benchmarks, what are their qualifications?

The auditors are individuals certified by the TPC. A current list can be found at www.tpc.org. They have years of experience in implementation of transaction processing systems, auditing of performance benchmarks and auditing of TPC benchmarks. Also, specific auditors are certified for specific TPC benchmarks.

9.3 How are Auditors approved by the TPC?

The auditors are screened by the Steering Committee, pass a written test, serve an apprenticeship, are questioned by a qualified panel and voted for acceptance by the General Counsel.

9.4 How are Auditors compensated?

They are paid by the test sponsors.

9.5 How do Auditors decide complex or ‘gray’ areas where specifications might be ambiguous?

The auditors may confer among themselves or raise the issue on behalf of the test sponsor to the Technical Advisory Board.

9.6 What process do auditors follow?

Each TPC benchmark has a clause describing the auditing process for that benchmark. The auditing process for TPC-W is described in Clause 9 of the specification. The audit process is designed by the auditor for the purpose of validating all aspects of the benchmark implementation and the published performance result.

9.7 What recourse do competitors have if they feel an Audit was not done properly?

They may raise the issue to the Technical Advisory Board if it is a technical issue or the Steering Committee if it is an ethical issue.