



# Performance Benchmarks

Benchmarks are a series of tests that are run on a computer or database so customers can compare vendor products. How important are benchmarks in today's computer market? Customers make purchasing decisions based on the results and vendors design and tune their machines to perform better in benchmark tests. This white paper will focus on TPC tests, but also discuss NetBench and ServerBench benchmarks.

The Transaction Processing Performance Council (TPC) is the most well known benchmark organization. The TPC is a democratic organization founded in the early 1980s and is made up of individuals from a wide range of vendors. They have cooperatively designed a series of tests to measure power, throughput and price/performance for computer systems in order to disseminate objective, verifiable data to the industry. Here is a summary of the TPC benchmarks:

- TPC-A - measures OLTP transactions and is seldom used today.
- TPC-B - measures database stress and is seldom used today.
- TPC-C - measures complex OLTP transactions and is used extensively today.
- TPC-D - measures Decision Support (DSS) and is used extensively.

The TPC-A benchmark was issued in November 1989, and measures performance in update-intensive database environments for OLTP (on-line transaction processing) applications. TPC-A measures the number of transactions per second a system can perform when driven from multiple terminals.

The TPC-B benchmark was issued in August 1990, and is designed as a database stress test. It tests disk input/output, system and application execution time, and transaction integrity. TPC-B measures throughput in transactions per second.

The TPC-C benchmark was issued in July 1992, and is designed to measure on-line transaction processing (OLTP). This benchmark is used widely in today's computer environment. TPC-C benchmarks are different than TPC-A benchmarks in that they are considered "complex queries". TPC-C measures multiple transaction types, more complex database and overall execution structure. TPC-C involves a mix of five concurrent transactions and is measured in transactions per minute (TPM). These tests are designed to test OLTP applications that manage, sell, or distribute a product or service.

The TPC-D benchmark was issued in April 1995, and is designed to measure Decision Support Systems (DSS). The TPC-D benchmark is used widely in today's DSS environment and measures applications that require complex, long running queries against large complex databases. There are 17 complex query tests including two update tests. The council sets data volume points of 1GB, 10GBs, 30GBs, 100GBs, 300GBs, 1TB, 3TBs, and 10TBs. Each volume point is considered a different benchmark.

## TPC-C Benchmark

The TPC-C benchmarks test complex queries in On-line Transaction Processing (OLTP). TPC-C involves a mix of five concurrent transactions and is measured in transactions per minute (TPM). These tests are designed to test OLTP applications that manage, sell, or distribute a product or service. OLTP Systems and each audited TPC-C result provide a published three-page executive summary and a full-disclosure report. The full-disclosure report includes details on system configuration, system cost, and benchmark methodology.

### **The primary metrics are:**

- **TPC-C new-order transaction rate:** There are five transactions that make up the transaction rate measured as transactions per minute (tpmC). The transactions operate against a database of nine tables. Transactions do update, insert, delete, and abort on primary and secondary key access. The five transactions consist of:
  1. **New-order:** New orders for customers are accepted.
  2. **Payment:** Updates to customer balance reflect a payment transaction.
  3. **Delivery:** Delivery orders are done as batch transactions.
  4. **Order-status:** Obtains status of customer's most recent order.
  5. **Stock-level:** This transaction monitors warehouse inventory.

- **TPC-C price/performance:** The price/performance is derived by taking the price of the entire system and dividing by the performance which equals price per tpmC. The price/performance numbers are often mistaken for the cost of the computer or host machine. This is not true! TPC benchmarks are system-wide benchmarks, encompassing almost all cost dimensions of an entire system environment the user might purchase, including:
  - Terminals
  - Communication equipment
  - Software transaction monitors and database software
  - Computer system or host
  - Backup storage
  - Three years maintenance costs

## TPC-D Benchmark

The TPC-D benchmarks test Decision Support Systems (DSS). Each audited TPC-D result provides a published three-page executive summary (contains three primary and two secondary metrics) and a full-disclosure report.

### **The primary metrics are:**

- **TPC-D power:** This metric is a query-per-hour rate for a single user. The user runs 17 varying queries and two update functions in a serial fashion.
- **TPC-D throughput:** This metric represents a query-per-hour rate for multiple users. Concurrent execution of 17 queries and two update functions are run to measure the total work supportable by the system.
- **TPC-D price/performance:** This price/performance ratio metric identifies the 5-year cost of the system in relation to the query-per-hour rating of that system.

### **The two secondary metrics are:**

- **Load time:** This metric represents the time it took to load the test data from a client or mainframe. There are ways to make load times appear faster and easier.
- **Disk storage ratio:** This metric describes the ratio of volume point raw data and the actual total disk space used in the benchmark. This ranges from 2.9 at the low end to much higher. Excessive indexing causes the higher volume points.

## What TPC-Ds Do NOT Measure

- **Ad hoc queries:** These are spontaneous queries that are found in real world data warehousing sites. Most (DSS) systems don't know tomorrow's questions.
- **Multi-user abilities:** A system can be tuned to run well with one user, but may perform poorly with multiple users. Compare single-user and multi-user averages.
- **Complexity of database setup:** Because TPC-D benchmarks use pre-defined queries some vendors create synthetic indexing designed to enhance results.
- **Performance on customer applications:** A customer should look at all 17 queries, but focus on the execution times of queries they expect to find in their environment. Customers can run benchmarks using their data and applications.
- **Why some vendors aren't published:** If a data warehouse vendor has not published TPC-Ds at a certain volume point they usually can't!
- **Comparisons to different volume point results:** A TPC-D benchmark at 300 gigabytes should be reasonably comparable with a benchmark at 100 gigabytes to show linear scalability.

## ServerBench

ServerBench is a Ziff-Davis benchmark which measures the performance of a server in a client/server environment. ServerBench produces numerous results, but its' primary result is an overall score for a server. The ServerBench setup places data and applications on the server with the client PCs used as front-ends to provide access to the applications. When using ServerBench the clients and the controller must be running Windows 95 or Windows NT. The ServerBench test environments include:

- The server you are testing
- PC clients
- One PC designated as the controller

ServerBench differs from NetBench in that ServerBench measures the performance of application servers, while NetBench measures the performance of file servers.

## **NetBench**

NetBench is a Ziff-Davis portable benchmark which can be loaded locally from disks on system which you wish to test. It measures how well a file server handles file I/O requests from as many as four different client operating systems. The file types include: DOS, 32-bit Windows, 16-bit windows, and or Mac OS systems. The clients access the server with requests for network file operations. Each client keeps track of how many bytes of data it moves to and from the server and how long the process takes. The information is used by the client to calculate its throughput for that test mix. NetBench totals all the client throughputs together to produce the overall throughput for a server.

NetBench differs from ServerBench in that NetBench measures the performance of file servers, while ServerBench measures the performance of application servers.

## **SAP Benchmark**

The SAP benchmarks measure server products used for running the SAP R/3 software application environment. The benchmark is designed to measure performance of a specific client/server system configuration combining server(s), relational database and SAP R/3 software.

### **The primary metrics measured are:**

- Number of benchmark users.
- Average dialog response time (measured in seconds) which must be less than 2 seconds for the results to be valid.
- Configuration throughput - measured in SAPS (the number of dialog steps per hour).
- R/3 module type (read on for types).

SAP measures database performance differently than TPC-C and TPC-D. The TPC-C and TPC-D benchmarks compare different databases that are accessed by SQL. This does not always reflect the performance the user could expect in real world applications. This is because most users do not use an isolated database system, but instead use an application system in which the system is an integrated component. The data is not accessed directly via the database system's interface (e.g., SQL) but rather via the predefined interfaces of the application system.

Currently, there are seven different SAP application benchmark modules available. The benchmarks are designed to perform the most typical transactions of the business model reflected by the application model. The benchmarks measure all performance relevant parameters such as database request times, wait times, CPU utilization, average dialog response by a given number of benchmark users and the achieved throughput. The seven benchmarks available are:

## *P e r f o r m a n c e   B e n c h m a r k s*

- **SD Benchmark** - Measures Sales and Distribution by creating orders and deliveries for orders. It also displays and changes the delivery and post goods.
- **FL Benchmark** - Consists of 4 customer postings and the line item display of the last posting. It also displays the open items.
- **MM Benchmark** - The Materials Management benchmark creates a purchase requisition for 5 materials, and then a purchase order with reference to the purchase requisition.
- **PP Benchmark** - The Production Planning benchmark consists of transactions to create a production order, change the order, and create completion confirmations.
- **WM Benchmark** - The Warehouse Management benchmark creates a transfer requirement to put 3 materials on stock.
- **PS Benchmark** - The Project System benchmark creates a project using a project profile, execute the cost planning for the project, and approve the project.
- **HR Benchmark** - This is a batch-mode background benchmark.

## Key Points To Remember

- Benchmarks are a series of tests which measure power, throughput and price/performance of computer systems and software.
- TPC-C benchmarks are used primarily for measuring complex OLTP systems.
- TPC-D benchmarks are used primarily for measuring Decision Support Systems.
- Vendors and customers need to take benchmark tests with a few grains of salt. Often the manufacturers of the products being tested will spend months tuning and testing their products to get the best results possible. These results often do not reflect real world experiences or acceptable practices.