Gartner
RPE2 Methodology
Overview

Gartner.

Contents

Gartner
RPE2 Methodology
Overview

Among the Gartner research and advisory products are tools to help clients understand and compare various characteristics of server systems, including pricing, features and performance. Our server performance estimates are powered by a methodology called Relative Performance Estimate 2 (RPE2).

RPE2 provides rapid approximate assessments of relative server performance. These performance estimates aid in comparing servers, making server recommendations or purchase decisions, analyzing server consolidation and technology refresh scenarios, capacity planning, and defining chargeback valuations.

This overview serves as an important reference for users of RPE2. It explains how the performance estimates are derived and how they can be utilized.

Ideas International was acquired in 2012 by Gartner, Inc. The acquisition includes proprietary methodologies such as RPE2.

Introduction

Comparing the performance of computer systems can be a difficult task. Unlike most other high-technology and mass consumer products, servers have little objective information available to help users quantify performance capabilities in absolute or relative terms. Server manufacturers are not obligated to provide standardized performance data on products, as they are with power characteristics.

In the absence of mandatory standards for specifying server performance, the IT industry has evolved a number of performance benchmarks. Some of these benchmarks are regulated by industry consortiums, others are created by individual hardware or software vendors, and still others are created by third-party companies or individuals. Benchmarks vary in terms of the workloads simulated, the runtime complexity involved, the system characteristics measured and the rules applied for testing and reporting results.

Users who require performance estimates often find available benchmark results inadequate. Common reasons for this inadequacy include:
- Data is not available for the specific product(s) users wish to evaluate.
- Data that is available focuses on only a subset of system performance or does not reflect the workload(s) users intend to run.
- Data is only available from a source with a possible conflict of interest, such as a seller promoting the purchase of a specific product.

The chances that relevant benchmark data will not be found are greater when users are trying to compare servers from different manufacturers, different architectures or different generations.

In the absence of comprehensive, relevant and comparable data from manufacturers, we created a theoretical performance estimate called RPE2. The objective of RPE2 is to provide users with comparable performance information for server products. RPE2 accomplishes this by incorporating the following:
- A composite workload profile
- Coverage of all x86, IA-64 and RISC server variants from the leading global manufacturers
- Coverage of current and obsolete server models

RPE2 is a theoretical performance estimate and not an actual observed measurement of server performance. It is largely based on published benchmark results and relative performance ratings from server manufacturers.

Like any single metric, RPE2 can provide useful information, but it can also be misapplied. Later in this document we will be discussing recommended ways of using the relative performance data as support for server selection or for various aspects of operational planning.

RPE2 Background

The original performance ranking data for enabling architecture-independent server comparisons was introduced in the late 1990s. This ranking series, known as RPE, was based purely on a single lightweight online transaction processing (OLTP) workload.

The focus on a single benchmark workload presented a number of limitations and risks. So in order to better represent and expand the wide breadth of workloads and software stacks being deployed on servers and to increase the pool of reference benchmark data points, a composite benchmark methodology was initiated. In 2005, the new enhanced ranking methodology, RPE2, was introduced as the replacement performance ranking series for server comparisons and consolidation tools.

RPE2 Definition

RPE2 is a composite benchmark, meaning that server performance characteristics are captured and calibrated against multiple workload profiles represented by a mix of industry benchmarks that have the widest technology coverage. The published or estimated performance points for each server processor option are aggregated by calculating a geometric mean value. In the standard RPE2, all components are weighted equally to prevent RPE2 from skewing toward a single benchmark or workload type. Other weighting options are described below.

A composite mix benchmark offers the following advantages:
• The multiple components represent a broader range of workloads and server architecture characteristics.
• Multiple components enable the impact of benchmark life cycles to be managed in a less disruptive manner; benchmark substitution can be handled within the existing framework, and the overall spectrum of results can be kept broadly consistent.
• Multiple components increase the likelihood that more absolute performance values contribute directly to the composite.
• Multiple components enable the incorporation of additional components and mitigate the enforced loss of a single component.

The initial RPE2 benchmark set was selected from all available industry and ISV benchmarks based on how complete their coverage was for the major manufacturers and server architectures, and the extent of their published results. The current RPE2 set includes the following six benchmark inputs in its calculation: SAP SD Two-Tier, TPC-C, TPC-H, SPECjbb2015, and two SPEC CPU2006 components.

As additional performance points for missing technologies appear in other existing benchmarks, or if new industry benchmarks are developed that potentially satisfy our selection criteria, they will also be considered for inclusion within the RPE2 composite.

The primary objective for RPE2 is to reflect benchmarked server family relationships and vendor ranking data on a benchmark component-by-component basis. The RPE2 calculation process extrapolates and interpolates the best-case performance data from multiple benchmark sources, including relative performance data provided by server manufacturers. By representing a broader spectrum of measured outcomes, the RPE2 values are more representative of overall server capability and the range of applications now being consolidated on virtualized server environments.

The RPE2 performance rankings are designed to provide users with the most accurate and comprehensive coverage of available server performance. RPE2 data covers all processor configuration options for x86, IA-64 and RISC servers for major vendors from 1997 onward — over 33,000 configurations in all.

RPE2 also has several other virtues not found in any other performance-ranking data:
- **Independence.** RPE2 was developed and is maintained by an independent analyst company.
- **Transparency.** RPE2 is the only composite benchmark mix that is fully documented.
- **Comprehensiveness.** RPE2 is the only composite ranking that incorporates multiple workload types and covers all major technologies and architectures, including virtual machine types and public cloud servers.
- **Durability.** RPE2 is designed to adapt to the availability and life cycle of the individual benchmark components; RPE2 continues even when constituent benchmarks change, become obsolete or are replaced.

RPE2 Workload Extensions

In 2010, RPE2 was expanded to include Workload Extensions. Workload Extensions use different weightings of the constituent benchmark components of RPE2 in order to highlight performance within specific workload profiles. The following RPE2 Workload Extensions were created:

- RPE2-ERP, highlighting the SAP SD Two-Tier component
- RPE2-Java, highlighting the SPECjbb2015 component
- RPE2-OLTP, highlighting the TPC-C component
- RPE2-Compute-Intensive, highlighting the SPEC CPU2006 components

While RPE2 provides a readily accessible, inclusive and convenient measure of server performance, quantitative benchmarking of specific workloads on target servers will likely provide higher accuracy than RPE2 in terms of predicting performance. We recommend that, when they exist, users should utilize workload- or application-specific performance benchmarks over RPE2 to estimate the performance of those workloads on specific servers.

Before reviewing RPE2 usage in more detail, RPE2 should be placed into the context of its primary sources. Because RPE2 values are mainly derived from benchmarks and other information provided by manufacturers, the caveats that the manufacturers apply to their own performance data must also apply to RPE2. Therefore, the following cautionary statements apply to RPE2:

• The source data may include manufacturer performance rankings, which are largely based on estimates rather than actual performance measurements.

• RPE2 and RPE2 Workload Extensions are based on benchmarks that include a mix of workloads with different characteristics, and these may not be representative of a user's intended workload.

• The published benchmark results used in the calculation of RPE2 are generally tuned by manufacturers for maximum possible performance, and, therefore, they may be dependent on specific hardware configurations, software packages and/or kernel, middleware and database settings in the software stack.

The following sections review each of the RPE2 usage scenarios in more detail. Note that Gartner also provides tools to undertake many of these specific tasks.

Server Consolidation Projects

Typically there are four major practical problems with server consolidation projects:

1. The number of servers needing to be analyzed and/or replaced is overwhelmingly large — perhaps many hundreds or thousands.

2. The environment is likely to consist of a wide mix of products with different architectures from various manufacturers, making objective, relative comparisons difficult.

3. Many installed servers are likely to be older or obsolete models for which manufacturer product lists or server management databases have no information.

4. Consolidation projects must be financially justified up front; quick rough estimates of scope and cost are needed before the time or money is available to perform in-depth analysis and consolidation planning.

In light of these challenges, how do you estimate how many servers can be consolidated on a given platform, and/or which servers are appropriate targets for consolidating a specific inventory?

We offer a solution that addresses these questions: the online Server Consolidation tool, alternatively called ServerCAR or the Server Consolidation module. The Server Consolidation online tool covers a very comprehensive list of servers (over 33,000 entries), along with their associated RPE2 values and environmental profiles, dating from 1997 onward.

Server Consolidation enables users to calculate the aggregate performance, memory and environmental details (such as power, heat output and rack space) of installed servers and then determine replacement options irrespective of the server technologies involved.

Server Purchase Assessments

RPE2 data is utilized by the Gartner Technology Planner and the IDEAS Competitive Profiles information services. These services offer comprehensive coverage of the comparative features, pricing and performance characteristics of a wide range of enterprise servers sold across major global markets. They use RPE2 data to identify which servers are likely to compete with each other based on their performance profile. This capability can be useful as a quick server shortlisting function.

Filling in Benchmark Gaps

Probably the most effective use of RPE2 data is for filling in gaps in benchmark coverage. A user will often know which benchmark is a good surrogate for his or her workload. However, the result pool for that benchmark may not include data on the specific server model the user intends to buy (or has already purchased). In this instance, adjusting the benchmark result of a tested model by the RPE2 ratio of the target to the tested model will give a good estimate of the target absolute benchmark value. This process assumes that the two models (tested and target) are from the same server family and are reasonably close in age and chip or core count. The further apart the models are within the range ranking table, the less reliable the estimated outcome will be.

Extrapolations of this nature can best be used as part of purchase assessments, as described previously, or for making capacity planning estimates in the absence of application-specific sizing tools.

Chargeback

RPE2 data can be used as an independent contributor to the establishment of chargeback rates. In such instances, this data can serve as a useful way of converting metered computing resources on various platforms into a harmonized charge-out algorithm.

Conclusion

While many server performance benchmarks exist, they frequently offer insufficient or inappropriate data for making sound assessments of relative server performance. RPE2 provides an accurate, independent and quick source of performance estimates to address this issue. RPE2 was developed to fill the extensive vacuum of missing performance information left by the server manufacturers. When used in conjunction with other Gartner services, RPE2 data effectively supplements industry benchmark information and serves as a productivity aid for making rapid approximate assessments.

Industry
Benchmarks

The following provides some background on the industry-standard benchmarks referenced in this overview.

Transaction Processing Performance Council (TPC)

The TPC is a nonprofit organization that develops transaction-processing and database benchmarks. Generally, TPC benchmark results must be audited before they become official, and TPC benchmarks are among the few that require full disclosure of system price along with performance. Gartner is an associate member of the TPC. For more information, consult the TPC website at http://www.tpc.org.

TPC-C. First released in 1992, this OLTP benchmark has been run on servers of many different architectures. The TPC-C test suite simulates an online order-processing system, executing transactions such as storing orders into a database and checking the status of existing orders. This benchmark tests not only the compute capabilities of the processor, but also the performance and capacity of memory and I/O. The longevity of TPC-C makes it useful for comparing the performance of older servers against the performance of current systems.

TPC-H. TPC-H is an ad hoc decision support benchmark that simulates complex search queries presented to a large database. To account for performance variances associated with the data-set size, TPC-H can be run across six different scales (database sizes) that range from 100 GB to 30 TB. Both the hardware configured and the database software deployed are significant contributors to TPC-H performance.

Standard Performance Evaluation Corporation (SPEC)

The nonprofit SPEC develops a number of benchmark suites that test specific subsystems of server hardware and software. Gartner is a member of the Open Systems Group (OSG) of SPEC. For more information, consult the SPEC website at http://www.spec.org.

SPEC CPU2006. SPEC CPU2006 runs compute-intensive workloads and consists of two benchmark suites: "INT" for measuring compute-intensive integer performance, and "FP" for measuring compute-intensive floating-point performance. The data used by the floating-point tests generally does not completely fit within internal processor caches; thus, these tests also stress the processor's cache/memory hierarchy. The SPEC CPU2006 suite replaced an earlier CPU2000 suite, which itself was the successor of an earlier SPEC CPU benchmark.

SPECjbb2015. SPECjbb benchmarks evaluate the performance of servers running typical Java business applications, as well as aspects of the Java virtual machine (JVM). This benchmark simulates order processing and data mining for a supermarket company, including response-time requirements on transactions. In addition to testing the processor and memory, these benchmarks also stress the operating system and JVM components.

SAP Standard Application Benchmarks

Unlike the benchmarks from SPEC and the TPC, the SAP benchmarks are controlled by a single company (SAP) and its partners, rather than an industry consortium. These benchmarks are designed to test the hardware and database performance of SAP applications and components. For more information, consult the SAP benchmarks website at http://www.sap.com/solutions/benchmark/index.epx.

SAP Sales and Distribution (SD) Two-Tier. This benchmark quantifies the performance of the SAP SD application, one of the many SAP application solutions. It measures a system with a database, server(s) and some form of SAP's enterprise resource planning (ERP) software.

About
Gartner

Gartner, Inc. (NYSE: IT) is the world's leading information technology research and advisory company. We deliver the technology-related insight necessary for our clients to make the right decisions, every day. From CIOs and senior IT leaders in corporations and government agencies, to business leaders in high-tech and telecom enterprises and professional services firms, to technology investors, we are the valuable partner to clients in 12,000 distinct organizations. Through the resources of Gartner Research, Gartner Executive Programs, Gartner Consulting and Gartner Events, we work with every client to research, analyze and interpret the business of IT within the context of their individual role. Founded in 1979, Gartner is headquartered in Stamford, Connecticut, U.S.A., and has 5,200 associates, including 1,280 research analysts and consultants, and clients in 85 countries.

Corporate Headquarters
56 Top Gallant Road
Stamford, CT 06902-7700
U.S.A.
+1 203 964 0096

Europe Headquarters
Tamesis
The Glanty
Egham
Surrey, TW20 9AW
UNITED KINGDOM
+44 1784 431611

Asia/Pacific Headquarters
Gartner Australasia Pty. Ltd.
Level 9, 141 Walker Street
North Sydney
New South Wales 2060
AUSTRALIA
+61 2 9459 4600

Japan Headquarters
Gartner Japan, Ltd.
Atago Green Hills MORI Tower 5F
2-5-1 Atago, Minato-ku
Tokyo 105-6205
+81 3 6430 1800 JAPAN
+81 3 3481 3670

Latin America Headquarters
Gartner do Brasil S/C Ltda
Av. Das Nações Unidas 12551, 25
Unit 2501 A
São Paulo 04578-903
BRAZIL
+55 11 3043 7544

Gartner