



IDEAS INTERNATIONAL ADVISORY PAPER: RECOMMENDED USAGE OF IDEAS RELATIVE PERFORMANCE ESTIMATES

October 2009

SERVICES

- » Server CAR
- » CPSystems
- » CPService

TABLE OF CONTENTS

Executive Summary	1
Introduction.....	1
RPE2 Background	2
Usage of RPE2	4
Server Consolidation Projects	4
Server Purchase Assessments.....	5
Filling in Benchmark Gaps	5
Charge-Out.....	6
Conclusion	6
Data Licensing and Subscription	6

Executive Summary

One of the important considerations when selecting a new server is whether it will offer adequate performance for its anticipated workloads. While there are numerous decision factors to consider when evaluating server solutions, each of the potential solutions must be able to meet the performance requirements.

Among the Ideas International (IDEAS) industry research and advisory products are tools to help customers understand various characteristics of the systems they may be considering, including pricing, features, and performance. To assist customers in narrowing the choices among the range of available server configurations, IDEAS developed a performance estimating methodology that helps identify server configurations with similar performance characteristics.

The IDEAS performance and pricing research tools are primarily designed as productivity aids to assist with rapid ballpark assessments, especially in areas where there is insufficient industry performance data. The IDEAS performance data is most appropriate for initial guidance and not for providing definitive answers. While IDEAS considers the performance data a valuable aid to the initial evaluation and short-listing process, IDEAS does not recommend using the data as the deciding factor in choosing one system over another. If performance is key to any final decision, then other performance data, such as actual workload benchmarking, should be used.

This white paper serves as an important background reference for anyone who intends to make use of Ideas International's relative performance estimates. It explains the derivation of the estimates and how they are best used.

Introduction

The evaluation of comparative computer systems performance can be a difficult task. Unlike most other high technology products and mass consumer products, there is very little objective information available to help the user quantify the capabilities of the server in absolute or relative terms. In fact there isn't a single attribute, apart from environmental information, that manufacturers are obliged to specify to help define or position their individual products.

By comparison, a car buyer is swamped by mandatory or voluntarily supplied specifications on all aspects of vehicle performance – covering top speed, acceleration times, engine power, fuel consumption, emissions, crash protection, etc. Some of these specifications are inherent; some are derived from controlled environment testing; others are the result of testing by reviewers. The specifications for every model have to show a minimum industry- or government-approved level of detail. Some, but not all, models are thoroughly reviewed, however rarely is any attempt made to simulate real world usage. Purchasers may enjoy reading the performance specifications, but they probably don't believe they will be capable of reproducing the results. Realistically, such information only gives buyers some guidance on which models meet their criteria. Buyers also understand that higher ratings or excellence in one performance attribute is not always a predictor of excellence in another.

How does all this compare with the computer market? In the absence of mandatory IT industry standards for defining and measuring performance, the industry has evolved a number of

INDUSTRY BENCHMARKS

There are many benchmarks that are used in the industry. We have included here just a brief description of those benchmarks relevant to our discussion. If you want more information on a given benchmark, we have included a link to the website of the organization that owns the benchmark. Ideas International also maintains a public website that summarizes many benchmark results in a convenient place. Go to <http://www.ideasinternational.com/benchmark/bench.html>.

TPC-C

First released in 1992, this OLTP benchmark has been widely measured on many systems. The longevity helps to equate performance on older, outdated servers with the performance offered by current systems – particularly useful for the CAR report. However, older benchmarks like TPC-C may no longer represent current computing techniques. As such, TPC-C is planned to be replaced by TPC-E (see below). The TPC-C test suite simulates an online order processing system, executing transactions like ordering, checking the status of an order, etc. This benchmark tests memory and I/O performance and capacity in addition to processor compute capabilities. This benchmark is one of the few that require reporting of price in addition to performance.

TPC-E

In 2007, the TPC issued a ratified a specification for this benchmark. As of May 2009, there are only twenty plus published results, primarily on x86 technologies. At this stage it is not possible to predict when the benchmark can be viewed as a full replacement for TPC-C.

TPC-H

This is an ad-hoc decision support benchmark. The Transaction Processing Council is the owner of the TPC benchmarks. The TPC members include most of the major hardware and software vendors in the industry. Ideas International is an associate member of the TPC. For more information on the above benchmarks you should consult the TPC website at <http://www.tpc.org/>.

performance “tests.” These benchmark tests are entirely voluntary and are uncoordinated with regard to ensuring that an overall representation of performance can be gained for an individual server, or even a server family. When benchmark tests are undertaken, the tests can be limited in scope, and therefore easy and inexpensive to run, or conversely, very complex requiring extensive and expensive tuning and measuring. The cheaper a benchmark is to run, the more results will be published, and vice versa.

The outcome is a major predicament for the server purchasers. Purchasers often find that the more comprehensive benchmarks don't cover the product(s) they are evaluating or even the style of workload they need to run. Hence, they must rely on the simplistic tests, which are more likely to include the product(s) they are evaluating.

In the absence of comprehensive relevant, pertinent, or comparable data from manufacturers, IDEAS created a theoretical performance estimate ranking called RPE2 (short for Relative Performance Estimate – version 2). The objective of RPE2 is to provide users with comparable performance information for server products during the initial evaluation phase. RPE2 does this by incorporating:

- » A composite workload profile
- » Full coverage of all x86/IA-64/RISC server variants from the leading global manufacturers
- » Coverage of current and obsolete models

The creation of the RPE2 rankings is entirely theoretical and is largely based on performance data from the manufacturers supplemented by published benchmark performance data.

While the RPE2 data uniquely fills many performance gaps, like any single specification value, it is capable of adding informatory value and being misinterpreted /misapplied.

This white paper discusses the recommended ways of using the IDEAS relative performance data as a contribution to server selection or various aspects of operational planning. RPE2, and its RPE predecessor, are proprietary information services of IDEAS and are licensed for internal use only by subscribers of IDEAS research offerings.

RPE2 Background

In the course of undertaking consultancy engagements, IDEAS faced the same difficulties as users in quantifying the performance of different servers. In the late 1990s, IDEAS developed an estimating methodology to add more rigor to the process of comparative server evaluations. Subsequently, the output of the methodology was named “Relative Performance Estimate,” or RPE. The original RPE was based purely on a lightweight Online Transaction Processing (OLTP) workload.

As such, RPE became constrained in two significant ways. First, all benchmarks have a finite lifecycle. A performance ranking calibrated against a single industry benchmark is exposed to the risk of that benchmark becoming discontinued, and there being insufficient and timely results from a valid replacement to take the calculation methodology forward.

Second, because the RPE output was based on a single benchmark, it was determining comparative rankings from a workload that, over the course of time, was becoming less representative of server capabilities across the industry and across the range of workloads and software stacks that users deploy on their servers.

In 2005, IDEAS decided to enhance RPE by creating a new ranking methodology – RPE2 – to overcome the previous constraints.

INDUSTRY BENCHMARKS (CONTINUED)

SPEC CPU2006 (SPECint_rate2006, SPECfp_rate2006)

The non-profit Standard Performance Evaluation Corporation (SPEC) develops a number of benchmark suites including CPU2006. SPEC CPU2006 is intended to compare 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. Note that the “fp” tests often do not fit within a processor chip’s internal cache and so the “fp” tests also stress the cache/memory hierarchy. The SPECint_rate2006 and SPECfp_rate2006 versions are used to measure multiprocessor server throughput.

SPECjbb2005

SPECjbb benchmarks evaluate the performance of servers running typical Java business applications as well as aspects of the Java Virtual Machine.

SPEC benchmarks are owned by the Standard Performance Evaluation Corporation. Most of the major vendors belong to the SPEC organization. You can get more information at its website: <http://www.spec.org/>.

SAP SD 2 Tier

This benchmark quantifies the performance of the SAP Sales and Distribution application, one of the many SAP application solutions. It measures a system with a database, server(s), and some form of the mySAP software. Unlike the benchmarks above, a software company rather than an industry consortium owns it. For more details on the SAP benchmark, please see <http://www.sap.com/solutions/benchmark/index.epx>.

RPE2 is a composite mix benchmark – meaning that server performance characteristics are captured and calibrated against multiple industry benchmarks. The published or estimated performance points for each server are aggregated by calculating a geometric mean value. No weightings are applied to the various components to prevent directly favoring any of the underlying performance components.

By working with a composite of a number of benchmarks, IDEAS met the following design objectives:

- » The multiple components represent a broader range of workloads and server architecture characteristics.
- » Multiple components enable the impact of benchmark lifecycles 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 extensive and complete the coverage was for the technologies and server architectures that IDEAS needed to assess. The current RPE2 set includes the following six benchmark inputs to its calculation – SAP SD 2-tier, TPC-C, TPC-H, SPECjbb2006, and two SPEC CPU2006 components.

As additional performance points appear in existing benchmarks for missing technologies, or new industry benchmarks are developed that potentially satisfy the IDEAS selection criteria, they will 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 from the best-case performance data from multiple benchmark sources. By representing a broader spectrum of measured outcomes, the RPE2 values are more representative of overall server capability and the range of applications that are now being consolidated on virtualized server environments.

The RPE2 performance rankings are designed to provide the best and most comprehensive coverage of server performance available to end users. RPE2 is the only performance ranking data that covers all x86 and RISC servers with their processor options from 1997 onward – 15000+ data points. It is therefore the only performance ranking data that is designed to be a user productivity aid for use in server consolidation studies.

RPE2 also has many other virtues that are not to be found in any other performance ranking data:

- » **Independence** – RPE2 was developed by an independent analyst company.
- » **Transparency** – RPE2 is the only composite benchmark mix that is fully documented. Similar vendor composite mixes are usually not available to end users and are never disclosed in content and methodology.
- » **Comprehensive** – RPE2 is the only composite ranking that incorporates multiple workload types and covers all major technologies and architectures.
- » **Open-Ended** – RPE2 is designed to adapt to the availability and lifecycle of the individual components. It continues even when components become obsolete or replaced.

Usage of RPE2

IDEAS RPE2 performance data can be used:

- » When you know the data you really need doesn't exist! Consolidation studies are a typical example. The RPE2 data is an invaluable default for getting some quantifiable information where none appears to exist.
- » When you want a quick assessment to get you into the right ballpark. Accuracy in the outcome is a low priority. If accuracy is important, the analysis should be largely supplemented by local measurement and use of suitable manufacturer sizing tools and guides.
- » When you know the characteristics of your workload can be represented by an industry benchmark, but the models you may need to evaluate are not featured in the results pool. In this case, the RPE2 values can be used to adjust or ratio known results against other possible server family alternatives.
- » When the benefit of a cross-technology ranking outweighs the need for accuracy. RPE2 can be used as a component of charge-out algorithms.

In any of the above circumstances, quantified information or guidance from vendors – which have greater insight into their own products or user requirements – should be given greater weighting or priority than IDEAS performance data.

Before reviewing RPE2 usage in more detail, the data should be placed into the context of its primary sources. Because RPE2 values are mainly derived from vendor performance rankings, the caveats that the vendors apply to their own performance rankings must also apply to RPE2. Typically, the cautionary statements will seek to distance any user interpretation or outcome from the supplied data for the following reasons:

- » The ranking data is largely estimated.
- » The ranking data is relative to a family model and is not an absolute measurement.
- » The ranking data is a mix of workloads with different characteristics that may not be representative or applicable to the users' intended workload.
- » The ranking data may be dependent on specific configuration content (i.e., maximum memory, or no disk activity, or lots of disk activity with an unconstrained I/O infrastructure).
- » The ranking data may be dependent on specific kernel, middleware, and database components in the software stack.
- » The ranking data may be dependent on the performance of specific third-party software products.
- » And, by implication, the ranking data may be dependent on the competence of their benchmarking teams!

Let's look at each of the RPE2 usage scenarios in more detail. IDEAS provides tools to undertake these and similar tasks in a number of research products.

Server Consolidation Projects

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

- » A huge number of servers need to be replaced – perhaps many hundreds or thousands.
- » The environment is likely to consist of a wide mix of products with different architectures from various manufacturers.

» The project team may not know exactly what workloads are being run and the utilization level of each server. Often there isn't even a proper description on the asset register as to what the servers are.

All that is known is that these workloads need to be moved onto replacement servers and that virtualization and partitioning may enable a large number of these older servers to be safely consolidated onto a single server or a smaller number of servers. How does one estimate how many servers can be consolidated on a given platform?

IDEAS has a solution that addresses this need – the Server CAR (Consolidation Analysis Resource). The Server CAR online tool covers a very comprehensive list of servers (over fifteen thousand entries), dating from 1996 onward, along with their associated RPE2 values and environmental profiles. Server CAR enables users to calculate the aggregate performance and environmental details – such as power, heat output, rack space – of the installed servers and to determine replacement server options irrespective of the server technologies involved.

Great precision is not needed in this context since there is frequently inadequate knowledge of server utilization – a key factor in the assessment. Achieving an “order of magnitude” understanding of the capacity requirements is more important. Server CAR helps users make rapid assessments on the impact of alternative consolidation scenarios and helps with the sensitivity analysis of utilization assumptions.

Server Purchase Assessments

RPE2 data is utilized by the IDEAS Competitive Profiles (CP) global information service. CP offers comprehensive reference coverage of the comparative features, pricing, and performance characteristics of a wide range of enterprise servers sold across major global markets. CP uses RPE2 data “under the hood” 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 short-listing function.

Competitive Profiles pricing options enable users to modify server hardware configurations or to add on the price of a full storage infrastructure. These changes can be made within the CP web application and exported for further analysis in client mode.

Extracting pertinent actual benchmark results held within the CP benchmark database and using that information as the basis of a price/performance analysis can improve the rigor of the analysis. In fact, a single benchmark may be more representative of the user requirement than the RPE2 composite. Or the most relevant benchmark for evaluation may be one that is not part of the RPE2 composite. In these situations, the benchmark comparisons may be very different to the RPE2-based outcomes.

Filling in Benchmark Gaps

Probably the most effective use of the RPE2 data is in filling in gaps for the benchmark you know is a good surrogate for your workload, but does not include data in its results pool for the server model you are about to purchase or have already acquired. In this instance, adjusting the benchmark result of the 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 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.

Americas

Ideas International, Inc.
800 Westchester Avenue
Suite N337
Rye Brook, NY 10573-1354
USA
Tel + 1 914 937 4302
Fax +1 914 937 2485

**Asia/Pacific and Worldwide
Headquarters**

Ideas International Limited
Level 3
20 George Street
Hornsby, NSW, 2077
Australia
Tel +61 2 9472 7777
Fax +61 2 9472 7788

Europe, Middle East, Africa

Ideas International Europe
The Courtyard
Lombard Street
Abingdon, Oxon OX14 5BJ
United Kingdom
Tel + 44 (0) 1235 462 890
Fax + 44 (0) 1235 462 891

actionable intelligence

www.ideasinternational.com



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

Charge-Out

RPE2 data can be used as an independent contributor to the establishment of “charge-out” rates, providing operational flexibility in the movement of workloads from one server to another. In such instances, RPE2 data can serve as a useful way of converting metered computing resources on various platforms into a harmonized charge-out algorithm.

Conclusion

IDEAS recommends that users apply valid industry benchmarks, where they exist, as the first priority for competitive server family selection. Each hardware technology and its associated software stack has strengths and weaknesses for specific workload or application profiles. Frequently there will be insufficient or inappropriate benchmark data available to affect a sound competitive assessment. At this point, RPE2 become the pre-eminent source of ranking data and demonstrates its virtues by the breadth and depth of its coverage.

We have described how IDEAS has developed relative performance ranking data to fill the extensive vacuum of missing performance information left by the server manufacturers. We have shown that the RPE2 data can be used most effectively in conjunction with other IDEAS services to supplement industry benchmark information or as a productivity aid for rapid ballpark assessments.

In explaining how RPE2 is derived from vendor ranking data, we have emphasized that any vendor caveats must be applied equally to IDEAS data. Inherently, RPE2 data cannot be “better” than vendor data; it can only be more readily accessible, more inclusive by covering a range of vendors and technologies, and more convenient to use.

In using RPE2 data, users should recognize that they are benefiting from the convenience of its very comprehensive coverage and not from any inherent “accuracy” or superiority over actual tested results. IDEAS believes that RPE2 is a valuable tool that helps solve the problem of insufficient actual benchmark test results in the industry to allow many servers to be compared and contrasted. However, it is not a magic solution for comparing computer systems and it is best used for initial guidance and not for definitive answers.

Data Licensing and Subscription

Ideas International is the owner of the RPE2 and other related performance ranking estimates. All Relative Performance Estimate variants are proprietary data that has been developed for subscription customers. RPE2 values and comparisons are not to be used in the public domain or outside of the subscriber company. RPE2 values are internal-use only measures for valid subscribers to Ideas International’s research offerings.

If you want to talk to an analyst at IDEAS about gaining access to RPE2 and how to use it, please feel free to contact us. You can reach us by e-mailing rpe2info@ideasinternational.com. We will be happy to help you with resolving any questions that you might have.