GIGAOM RESEARCH

Discovering the best cloud for your applications through benchmarking

By Paul Miller

June 12, 2014

This report was underwritten by CliQr.

TABLE OF CONTENTS

Executive summary

For most businesses, the debate about whether to embrace the cloud is over. It is now a question of tactics — how, when, and what kind? Cloud computing increasingly forms an integral part of enterprise IT strategy, but the wide variation in enterprise requirements ensures plenty of scope for very different cloud services to coexist.

Today's enterprise cloud deployments will typically be hybridized, with applications and workloads running in a mix of different cloud environments. The rationale for those deployment decisions is based on a number of different considerations, including geography, certification, service level agreements, price, and performance.

Price and performance are often seen as closely – almost inextricably – linked, and this connection is only likely to grow stronger as basic cloud infrastructure becomes increasingly commoditized. For some workloads, cost may be an overriding factor in selecting a cloud provider, and for others, performance is all that matters. For most, a complex combination of these and other factors will lie behind any deployment decision, making it important to ensure that buyers are fully informed about their options.

Frequent high-profile price reductions, equally numerous but less visible infrastructure upgrades, and the less rapidly evolving landscape of certifications, accreditations, and legal considerations combine to ensure that even the best and most informed selection processes require regular review; the best cloud provider for a particular set of requirements today may be surpassed by a competitor tomorrow.

Key findings include:

- There is significant variation in the price and performance of competing cloud solutions, both public and private.
- Lack of consistency in specification or description makes it challenging to accurately compare the capabilities of competing solutions.
- The characteristics of specific applications (web servers, e-commerce applications, Hadoop clusters, etc.) mean that their performance will change from one cloud to another.
- The most effective way to really understand the best place to run a given application is therefore to test how that application performs in different clouds and to build a comprehensive view of the complex balance between price and performance for any given workload.

Cheap, fast, or other?

In high-frequency trading and financial fraud detection or the real-time placement of advertising, time really is money. Reaching an actionable decision just fractions of a second faster than a competitor can mean the difference between a successful profitable transaction and missing out. For companies like European advertising platform MobFox or fraud detection provider SQN Banking Systems, the speed with which their systems deliver accurate and actionable intelligence is a key differentiator, and they (and their customers) consider greater speed to be something worth paying for.

In areas such as video transcoding and genomics research, on the other hand, it is often more important to process very large volumes of data as cheaply as possible. Companies like video encoder/transcoder Ooyala and DNA sequencers DNAnexus manage costs by bidding on cut-price Spot Instances from Amazon Web Services (AWS) for much of their processing, often only running jobs when demand for AWS is low and the spot price for available computing resources falls below a predetermined figure.

Even at these extremes, companies make compromises. Most financial institutions may value and pay a premium for speed, but there are limits. Eventually, the additional cost required to achieve incremental increases in speed becomes prohibitive. In reality, they are buying the fastest service they can *afford* while remaining profitable, not necessarily the fastest service that is actually or theoretically *available* to them. At the other end of the scale, companies like Ooyala and DNAnexus need to get data processed. They are striking a balance between needing to rent computing capacity and wanting to pay as little as possible for that. If either company placed too much emphasis on cost-saving and set an unrealistically low cap on what they were prepared to pay, they would never get any work finished and would quickly go out of business.

All users of cloud computing infrastructure conduct a very similar cost-benefit analysis when they select a cloud provider. For some, this will be an explicit and conscious process of comparing competing cloud solutions, and their cost savings or performance gains in specific situations will be significant. For others, it will be a far more subjective undertaking, which runs the real risk of failing to identify the most effective cloud for their requirements.

HIPAA, data protection, and other considerations

There are, of course, other considerations than simply striking the right balance between price and performance:

- Regulated industries like financial services, education, and healthcare may need to find cloud providers with relevant accreditations (such as HIPAA for U.S. healthcare providers, PCI for payment processing, etc.).
- Companies processing personally identifiable information about EU citizens will need to select cloud providers that keep them in compliance with European countries' data protection legislation.

Although the relevant laws, audits, and compliance procedures may be complex, arcane, and (sometimes) associated with potentially punitive damages for breach, the impact upon selection of cloud providers is actually relatively light. A cloud provider either supports HIPAA-compliant workloads or does not. A specific data center either has ISO 27001 security certification or does not. A specific data center is either inside the European Economic Area or is not. Often the evaluation is simple and binary.

Increasingly the cloud providers themselves make it easy to discover whether or not they meet the particular audit, compliance, and legal requirements of a specific customer or industry. AWS, IBM, Microsoft, and others all provide explicit statements on their websites and keep these up to date.

Leveling the playing field

It can be extremely challenging to objectively evaluate the price and performance of competing offerings from cloud infrastructure providers like AWS, Google, or Microsoft.

Free, fee, or discounted?

Each cloud provider follows a different approach to charging for its products. Inbound and outbound networking costs, storage, compute, and a wide range of value-added services are typically billed separately. Each provider has its own basic price for each component, its own rules for increasing (or discounting) as volumes rise, and its own rationale for altering pricing through long-term subscriptions, bundling, spot markets, free offers, or targeted marketing campaigns.

All of this information is available, however, and can be collected and analyzed in order to reach an objective conclusion, particularly for non-compute aspects of the cloud provider's offering, such as network transfer and data storage costs. It would be unrealistic to hope to identify the cheapest (or most) expensive provider in abstract terms, but it is perfectly feasible to be more specific and to identify the cheapest cloud provider for storing 1TB of data for a month or the most cost-effective for uploading 100GB of new data per day for a year.

The price of these services can be discovered and compared, but performance is far less tangible. Understanding how *fast* data can be uploaded to a specific cloud or how *reliably* it can be stored there will require far more work. Matters become more complex when considering the aspect of a cloud that typically comprises the biggest piece of any monthly bill — compute.

Apples and oranges

Despite the efforts of companies like 6fusion, the cloud industry has no commonly accepted metric for describing a basic unit of compute. A machine with one virtual CPU and 8GB of memory from one provider may prove significantly faster than a superficially identical machine from a competitor, simply because one uses faster RAM than the other. To further complicate matters, most providers bundle their own combinations of CPU and memory to create products that are unlikely to be directly comparable with anything else. Rackspace's basic Performance 1 Flavor Class virtual machine, for example, includes 1GB of memory, 20GB of flash storage, and one virtual CPU. AWS' (approximately) similar m3.medium instance includes 3.75GB of memory, another single virtual CPU, and 4GB of flash storage. The details of memory speed, data throughput, processor performance, etc., are not provided by either company in a way that would enable easy comparison. The single virtual CPUs are unlikely to offer identical performance, and there will be few situations (if any) in which AWS' memory delivers 3.75 times the performance of Rackspace's.

Generic benchmarks, such as those recently conducted by InfoWorld, provide a degree of visibility into the performance that can be expected from different clouds for a small set of standardized workloads. Although they cannot provide insight into the likely performance of a specific customer workload, these benchmarks are certainly part of reaching an informed conclusion about the relative merits of competing clouds.

Products like RightScale's Cloud Analytics, Cloudyn, and Cloudability offer a degree of visibility on spending with competing cloud services. For those who have been running their own workloads in the cloud for some time, these products are also beginning to offer increasingly detailed advice about ways in

which existing spend can be optimized. All of these products look at spending on cloud infrastructure without providing a similar level of detail with regard to the performance of the resources being paid for.

Clouds are not uniform

Large multi-region clouds, such as those offered by the major cloud infrastructure providers, introduce additional complexity for benchmarking because of variation between a single provider's regions:

- Network latency will probably mean that a customer in New York City sees better performance from AWS' Virginia data center than from their Dublin facility, particularly for applications requiring significant data transfers.
- The age of physical hardware may hypothetically mean that a particular instance type performs marginally better in Rackspace's Sydney data center than in their London one.
- The specific hardware configurations in the data center may hypothetically mean that Google's European region is slightly better suited to high-volume data processing than their Asian region.
- Customer demand may hypothetically mean that Microsoft Azure's Singapore data center is more prone to "noisy neighbor" problems than their Illinois site.

All of these factors are subject to gradual change as providers upgrade hardware, as the mix of customers and workloads evolves, and as data center operators, customers, and intermediaries switch between network providers.

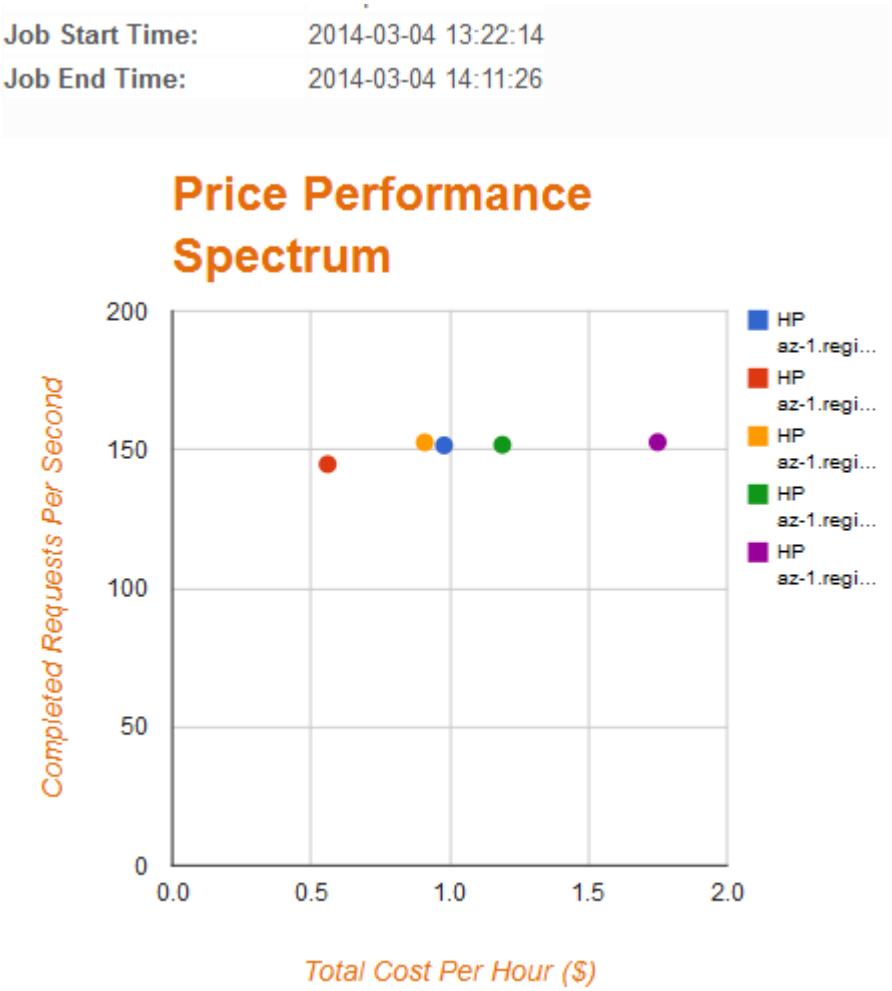
Benchmark applications, not clouds

Competing cloud infrastructure providers, as we have seen, offer products that can prove difficult to compare in a meaningful fashion. Furthermore, different regions within a single cloud may perform slightly differently to one another. The final complicating factors relate to the specific applications that a particular customer intends to run, as a cloud that excels in handling a moderately sized Hadoop cluster may perform less well than its competitors when tasked with running an e-commerce application or a complex website.

Customers with a keen interest in ensuring the best possible trade-off between price and performance are therefore left with only one realistic option: benchmarking the performance of their particular application in all of the regions operated by each cloud provider under consideration and factoring in all of the costs involved in using each competing cloud. This process could conceivably be completed manually, but there is a real risk that the cost and effort involved in creating and repeatedly running manual benchmarks will often exceed the value to be gained by using any findings to optimize cloud selection.

By automating the process of creating and repeatedly analyzing benchmarks to assess the price and performance of running specific applications on different cloud infrastructures, consumers of cloud gain the value of visibility and informed decision-making without the cost of manual benchmark management.

Figure 1. Price performance comparison for a generic Java web application

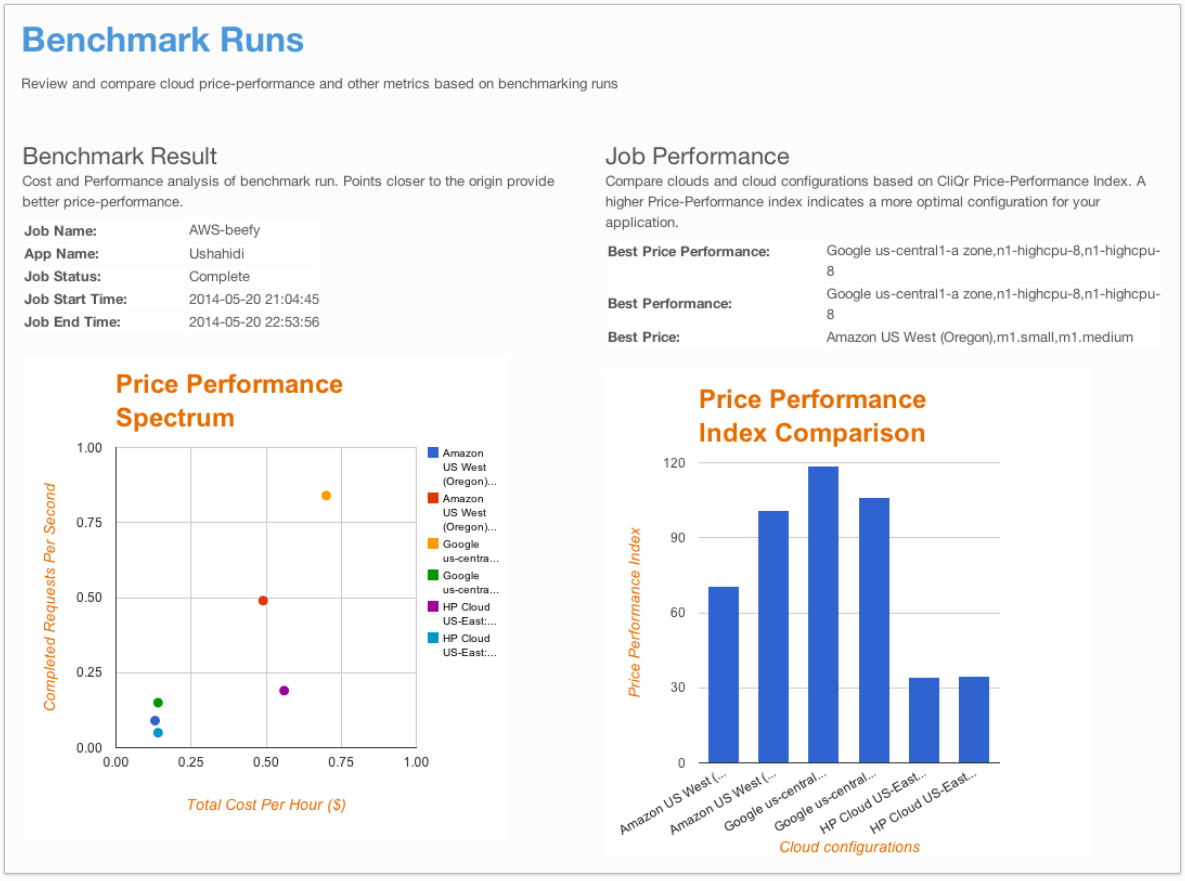


(Source: CliQr)

In this case, the application is being run in the U.S. central region of Hewlett Packard's public cloud, and a number of configurations are being tested to find the most beneficial for this particular workload. For this specific application, the figure clearly shows no value in paying a premium for the more powerful load balancers or application servers of the blue, green, and purple combinations. The additional cost of larger load balancers and application servers in the move from red to orange is offset by the use of a smaller database server. There *is* an improvement in performance, but it may not be sufficient to justify the associated increase in cost.

To provide a comprehensive picture, this benchmark would need to run in any HP Cloud region of interest to the customer and in the relevant regions of any other cloud provider also under consideration. A benchmarking tool simplifies the process but does not completely remove the need for manual intervention to specify applications, select clouds, and run the benchmarks themselves.

Figure 2. Comparison between three different cloud providers: AWS, Google Compute Engine, and HP's Helion public cloud



(Source: CliQr)

In each cloud, the Ushahidi application (and the Apache web server and MySQL database server it requires) is being tested on two types of virtual machine instance: the smallest (and cheapest) available to trial users of CliQr's benchmarking service and the largest (and most expensive). In this particular test, the U.S. east region of HP's public cloud (light blue for their smallest instance, purple for their biggest) consistently performs very poorly, and the U.S. central region of Google's cloud (green for their smallest instance, orange for their biggest) consistently performs well. The histogram also confirms that Google's biggest instance offers the best combination of price and performance for this particular workload. Other workloads or this workload run in other regions of these clouds might deliver very different results.

Benchmark, then switch?

One clear use case for this benchmarking is to find the best cloud to host a new application. A far larger opportunity is associated with moving existing workloads from one cloud to another as circumstances change.

In his keynote to the AWS Summit in San Francisco earlier this year, Amazon's Andy Jassy claimed the company cut prices for AWS 42 times in six years. Some of those cuts have been relatively minor. Others, such as those he announced in March, saw prices fall by as much as 60 percent in some areas. AWS is far from alone, with Google and Microsoft making similar sweeping cuts during March.

Price changes of this magnitude are likely to render any earlier benchmarking process void, requiring the customers who always want to be sure they're getting the very best performance for their money to run their benchmarks again.

Price certainly isn't the only consideration. Cloud providers change the specifications on their hardware from time to time or open new data centers that may be more accessible to customers. Any of these changes may be sufficient to make it worth running benchmarks again, and the new results will often suggest that it would be worth moving elsewhere. Those moves, with increasing complexity, might involve:

- Moving to a different instance or combination of instances in the same cloud region;
- Moving to a different region with the same cloud provider;
- Moving to a different cloud provider.

In each case, the expense, pain, and complexity of any move needs to be carefully weighed against the anticipated performance gains or cost savings of the new solution. Major public announcements around improved hardware or reduced pricing will often trigger a desire to re-benchmark. In the unlikely event that neither occurs for an extended period, it may still be worth re-running benchmarks every three to six months to test and reaffirm earlier decisions.

The benchmark says move; now what?

Certain workloads are extremely simple or generic in nature. These include:

- Modifying large sets of files with generic text processing and editing commands;
- Running simple scripts to convert large numbers of images from one format to another;
- Completing simple arithmetic operations at scale.

Workloads like these can run anywhere, and customers running them with enough regularity to justify benchmarking them in the first place may well realize compelling cost savings or performance gains by immediately moving to a new cloud whenever their benchmarks detect a beneficial change.

The next set of workloads is slightly more complex but still relatively easy to migrate. These include:

- Generic, self-contained, and well-understood applications, such as WordPress, Drupal, etc.

These workloads can often run anywhere too but carry the (slight) additional overhead of redirecting DNS entries, changing any relevant firewall rules, etc. There is likely to be value in moving these workloads in response to projected cost savings of more than a few percentage points.

A third set of workloads brings added complexity, probably reducing customer enthusiasm for a move unless the promised benefits are significant. These workloads include:

- Complex combinations of well-understood applications;
- Specific configurations of and connections between a number of different virtual machines;
- Applications that depend upon heavily customized or bespoke code.

Moving each of these introduces the risk that something will break or that apparently identical packages, libraries, and virtual machine types will not quite perform as expected. Customers are likely to be far

more wary about moving these applications and less interested in even considering the possibility unless the projected benefits are significant.

Solutions from companies like CliQr, Ravello, CloudSwitch (now part of Terremark), and CloudVelocity are specifically designed to ease the pain of cloud migration in cases such as these.

The fourth set of workloads is perhaps the most complex. While customers will normally be comfortable moving these from one region of a specific cloud to another, the complexity of moving to a new cloud will often be too high to contemplate. These workloads include:

- Tight dependencies to cloud-specific tools or technologies, such as AWS Elastic MapReduce, AWS Redshift, Google's BigTable, etc.

The future

Until recently, it has not been feasible to accurately benchmark the performance of applications running in different clouds. Adopters of cloud have therefore tended to develop their own criteria for assessing competing offerings, and they have also been reticent to switch providers too regularly. There is no real evidence to suggest that even the largest cloud price cuts have done much on their own to persuade existing cloud users to switch from one provider to another.

The availability of credible generic benchmarking solutions from companies like CloudHarmony and Cloud Spectator makes it possible for cloud users to gain a degree of visibility into the relative merits of competing infrastructure providers. CliQr is one company focusing this benchmarking on the specific application workloads their customers need to run. But for many customers in 2014, uncertainty, inertia, and the perception that most price cuts or performance boosts *tend* to be matched by competitors probably mean that benchmarking of this sort remains a luxury.

Today the real value proposition for cloud application benchmarking lies in serving the (small but growing) set of customers who:

- Use cloud computing resources at scale;
- Typically run fairly generic or simplistic workloads, which could run on any cloud;

- Operate with very tight margins in which small cost reductions or performance gains will have a disproportionately significant effect.

As the mechanics of application benchmarking become more accessible and as the migration tools become better at accurately and fault-tolerantly transplanting *any* workload from one cloud to another, the addressable market for these tools will quickly get much bigger. Rather than requiring customers to manually initiate a battery of benchmarks, for example, benchmarking providers might explore automatically benchmarking different classes of application, proactively notifying their customers when a migration might be practical.

Until then, the exercise of benchmarking remains one that is academically interesting and potentially enlightening but relatively rarely acted upon.

Key takeaways

- There is significant variation in the price and performance of competing public cloud solutions, and there may be some variation in the performance of notionally identical instances in different regions of the same cloud.
- Lack of consistency in specification or description makes it challenging to accurately compare the capabilities of competing cloud solutions.
- The characteristics of specific applications (web servers, e-commerce applications, Hadoop clusters, etc.) mean that their performance will change from one cloud to another.
- The most effective way to really understand the best place to run a given application is therefore to test how that application performs in different clouds.

About Paul Miller

Paul Miller is an analyst and consultant based in the East Yorkshire (UK) market town of Beverley and working with clients worldwide. He helps clients understand the opportunities (and pitfalls) around cloud computing, big data, and open data, as well as presenting, podcasting, and writing for a number of industry channels. His background includes public policy and standards roles, several years in senior management at a UK software company, and a Ph.D. in Archaeology.

Miller was Curator for GigaOM Pro's Infrastructure/Cloud Computing channel during 2011, routinely acts as Moderator for Gigaom Research webinars, and has authored a number of underwritten research papers such as this one.

About Gigaom Research

Gigaom Research gives you insider access to expert industry insights on emerging markets. Focused on delivering highly relevant and timely research to the people who need it most, our analysis, reports, and original research come from the most respected voices in the industry. Whether you're beginning to learn about a new market or are an industry insider, Gigaom Research addresses the need for relevant, illuminating insights into the industry's most dynamic markets.

Visit us at: research.gigaom.com.

© 2014 Giga Omni Media, Inc. All Rights Reserved.

This publication may be used only as expressly permitted by license from Gigaom and may not be accessed, used, copied, distributed, published, sold, publicly displayed, or otherwise exploited without the express prior written permission of Gigaom. For licensing information, please contact us.