CloudHarmony
trans parency for the cloud

State of the Cloud - Compute

16 July 2014

Version 1.7

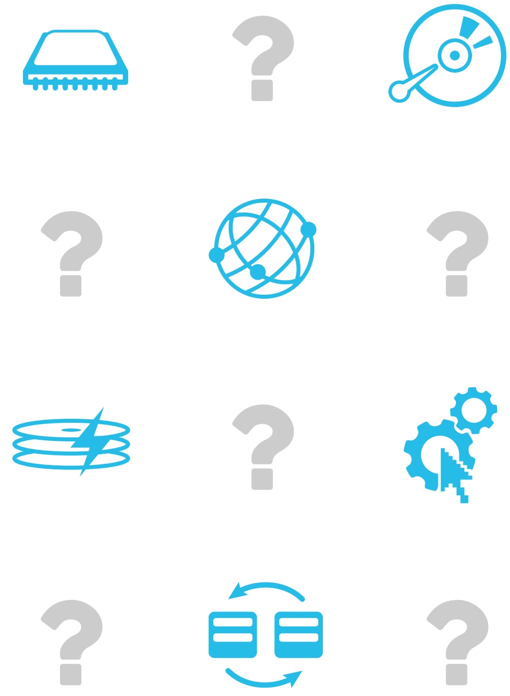
Abstract

This report examines how to use benchmarks to compare performance among cloud compute services. Actual comparisons of web and database server workloads are included as an example. Services covered in this report include Amazon Web Services (AWS), DigitalOcean, Google Cloud Platform, Microsoft Azure, Rackspace Cloud, and SoftLayer.

The purpose of this report is to help decision makers decipher often competing claims of performance superiority from vendors and reviewers, and to provide meaningful insight about cloud performance.

Introduction

If you're a decision maker who has attempted to compare the performance of cloud services, chances are you'll have encountered some or all of the following questions:



- Which cloud provider should I choose?
- How do I compare performance across cloud services?
- Is there a fastest cloud?
- Which provider is best for my workloads?
- Which provider offers the best value?

Comparison Challenges

You'll probably have found that it is nearly impossible to compare compute service performance based solely on information from providers themselves. For example, there is no standard for describing processing capacity of compute services. Cloud providers often use inconsistent and incomparable terms such as CPUs, vCPUs, cores, MHz, GHz, and more.

Most cloud compute services are virtualized. Virtualization lets providers operate more efficiently, provision compute instances faster, and offer lower costs to customers. Virtualization enables the partitioning of physical resources, such as CPU cores, to virtual resources, such as virtual CPUs (often using a one-to-multiple ratio). It also allows providers to oversubscribe resources, on the assumption that many users will not use all their resources all the time. However, virtualization doesn’t change the limited load capacity of the underlying hardware, and if providers do not properly manage virtualization, performance variability and unpredictability can result when many users try to use resources at the same time. Each provider uses a different virtualization strategy, with some using resource partitioning and oversubscription more aggressively in order to price services lower. Measuring and accounting for performance variability adds complexity to comparing cloud services.

Finally, it is difficult to find reliable comparisons of cloud services. While a number of studies have made information publicly available, many are of questionable quality because they are based on irrelevant benchmark selection or incorrect test execution, or they fail to incorporate cloud nuances into the analysis. Often these studies are used by companies or media to make broad claims and sensational headlines, adding confusion to the task of comparing services.

Optimal Benchmarking

This report is designed to help you filter through the noise and confusion by presenting a methodical and commonsense approach to testing and comparing performance of cloud services. Topics covered include:

- Selecting relevant benchmarks
- Accounting for workload-specific considerations
- Choosing comparable compute instances
- Running benchmarks and understanding results
- Non-performance related factors to consider

These topics are presented using a case study that compares multiple cloud compute services. The case study focuses on web and database server workloads, but the methodology is applicable to other workloads and cloud services. Future reports will cover additional workloads.

Providers in this Report

There are hundreds of cloud providers to choose from. In order to refine the scope of this report we selected six well-known providers, all of whom offer compute services in North America, Europe and Asia. Future reports will cover additional providers.

Amazon Web Services

Amazon Web Services (AWS) is a spinoff of online retailer Amazon Inc., based in Seattle, WA. AWS launched in 2006 as one of the first cloud infrastructure services, leveraging from lessons learned from the infrastructure and technology developed to maintain Amazon.com. Estimated earnings were $2.1 billion in 2012. Amazon is ranked #35 of Fortune 500 companies. AWS offers cloud services in ten regions, including four in the US, one in Europe, three in Asia, one in Australia, and one in Brazil. AWS provides a diverse range of services including compute, storage, security & access, networking & content delivery (CDN), databases (relational, NoSQL, columnar, and caching), analytics (managed hadoop, real-time streaming, and data warehousing), application services (queuing, orchestration, app streaming, transcoding, push notifications, email, and search), deployment and management (containers, DevOps, and templates), managed virtual desktops, and more. The AWS Partner Network (APN) includes nearly 7000 consulting/systems integrator partners and more than 3000 technology/ISV partners. AWS is a major player in the cloud space with hundreds of thousands of customers in 190 countries, more than 800 government agencies and 3,000 educational institutions. Key market differentiators include rapid innovation, aggressive pricing, large scale capacity, many high-profile clients such as Netflix, and a comprehensive suite of services.

DigitalOcean

DigitalOcean was founded in 2012 in conjunction with the startup accelerator TechStars. DigitalOcean provides compute instances starting at $5 per month. All compute instances run off local solid state disks (SSD). DigitalOcean seeks to gain market share in the low- to mid-range virtual machine (VM) hosting market through ease-of-use and aggressive pricing.

Google Compute Engine

Google entered the cloud space in April 2008 with the release of the Google App Engine (GAE) platform. In May 2010, Google released an object storage service and in June 2012, the compute service Google Compute Engine (GCE). Google was founded in 1998 and has more than 53,000 employees in more than 40 countries. Google reported revenues of $50.18 billion in 2012. Google is ranked #46 of Fortune 500 companies. Google refers to its suite of cloud services collectively as the Google Cloud Platform. GCE is currently available in US, EU, and Asia data centers and supports Linux-based compute instances only. GAE supports multiple programming stacks (including Python, Java, Go, PHP) that conform to specific runtime constraints. Key differentiators include Google's experience maintaining Internet scale infrastructure (estimated at hundreds of thousands of servers); a comprehensive suite of cloud services; many high-profile clients including Best Buy and Khan Academy; integration with related

Google services like Google Apps and authentication; and backing from an established, technology-driven Internet company.

Microsoft Azure

Microsoft entered cloud computing with the Microsoft Azure platform in 2008. Initially it was limited to .Net applications and object storage, but now includes support for SQL databases, CDN, and compute instances. Cloud computing is now a key strategic focus for Microsoft, and the company is investing heavily in this space. Microsoft was founded in 1975 and has more than 94,000 employees in 190 countries. It reported revenues of $73.7 billion in 2012. Microsoft Azure is available in eight regions including four in the US, two in Europe and two in Asia. Services include Windows and Linux compute instances, CDN, platform as a service (PaaS), object storage and more. Key differentiators include global infrastructure, a comprehensive suite of cloud services, support for Microsoft software, and backing of a Fortune 500 company (number 37).

Rackspace Cloud

Rackspace was founded in 1998 as a managed hosting provider and entered the cloud market in 2008 following the company's acquisition of Slicehost. Rackspace has more than 4800 employees and revenue for 2012 was $1.3 billion. Rackspace offers cloud services in multiple regions globally including three in the US, one in Europe, one in Asia and one in Australia. Services include Linux and Windows compute instances, object storage, CDN (using Akamai), DNS and more. Key differentiators include the company’s trademarked "fanatical support," upgrade path to managed services, and an established track record since 1998. Rackspace is also a strong supporter of the open-source cloud management platform OpenStack.

SoftLayer

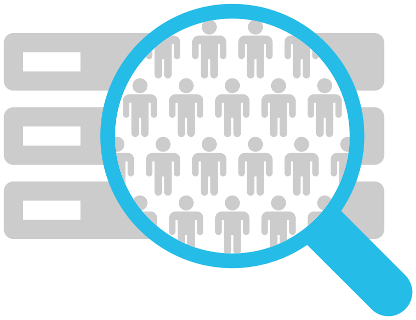
SoftLayer was founded in 2006 as a managed hosting provider. In 2010, SoftLayer merged with The Planet, and in 2013 it was acquired by IBM. SoftLayer offers both traditional managed hosting and cloud services. They own and operate multiple data centers globally, including four in the US, one in Europe and one in Asia. SoftLayer's cloud suite includes Linux and Windows compute, object storage, CDN, DNS, and more. Key differentiators include availability of both virtualized and bare metal cloud servers, being part of a Fortune 500 company (number 19), and offering both dedicated hardware and cloud services.

Cloud Benchmarking

Benchmarks have been used for measuring and comparing compute system performance for decades, and, used correctly, are also useful for evaluating cloud performance. When you run benchmarks on cloud systems, however, you must factor in subtle but significant differences to hardware-based systems, including multitenancy, oversubscription, and scalability. The benchmarks you select should also be reputable and relevant to your workloads.

Virtualization

Cloud compute services usually run in virtualized environments. As we discussed previously, virtualization allows providers to optimally utilize infrastructure through partitioning of physical resources such as CPU, memory and disk space, to virtual resources. Although virtualization is standard practice, if not properly managed, it can result in performance variability and inconsistency that can negatively affect your workloads.



Providers typically manage virtualization using policies that dictate resource partitioning, oversubscription, utilization limits, and actionable thresholds. For example, a provider might monitor to ensure that total CPU utilization on a physical server does not exceed 70 percent. These policies are usually undisclosed and vary between providers, which means you may need to investigate to determine how much variability to expect from a given service.

During benchmarking, if your results are inconsistent, this may be an indication that the provider is using more aggressive virtualization policies (or perhaps not properly monitoring utilization), and might be assigning too many users to the same resources.

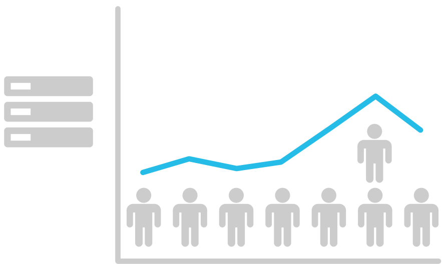
Price

Another telltale sign of oversubscription is price. Cloud providers have many common fixed costs. Power, for example, is a fixed cost for running hardware. A typical dual-processor server with 8-12 CPU cores uses about two to three amps of power, and a 20-amp power circuit might cost $300-600 per month in the United States. (Power costs vary depending on the locale.)

The cost of power for this server therefore might range between $30-90 per month. Adding bandwidth, cost of hardware, network equipment, spare capacity, staff, and other operational expenses could easily double or triple the monthly cost. Providers who operate their own data centers and at a large scale have more leverage to reduce costs, but probably only at a linear scale. Using this estimate for a single server with 12 physical CPU cores, the operational cost per core would fall between $8-23 per month. This means a provider offering virtual CPUs for less than $8-23 per month is likely oversubscribing the physical cores. For example, if virtual CPUs were priced at $5, the provider would need to oversubscribe by a ratio of two-to-five times to cover their costs, and because they likely also want to make a profit, oversubscription may be even higher.

Scalability

It is not especially difficult to start a cloud service. Simply purchase a server with the latest hardware, install free virtualization software, co-locate, and begin selling virtualized compute instances. Your first few users will likely experience excellent performance because they wouldn't be sharing resources with many users. Benchmarks could be used to show superior performance compared to established services. However, without additional servers to handle growth, performance will likely decline as more users sign up. You'll only be able to support a small number of users on your server, and you are not set up to scale.



It's important to have some sense of a provider's ability to scale and properly manage virtualization as they do. A few red flags to watch for are low resource limits (max compute instances per account), lack of automation, and long provisioning times. Also, you should take performance claims from new services with a grain of salt, because while there may be some truth to the claims now, their capacity to live up to those claims may diminish quickly as the service takes on actual customers.

Benchmark Selection

Benchmarks measure one or more performance characteristics. Hundreds of benchmarks have been created and used over the years. Some benchmarks measure very specific things while others are broad in scope.

Sometimes you'll see cloud performance comparisons based on a single benchmark, or based on benchmarks with limited relevance to actual workloads. In the past, flawed benchmark comparisons have been used to justify sweeping and sensational claims, such as *Provider X up to 14 times Faster than Provider Y*; or, *Provider X Named Fastest Cloud Service*.

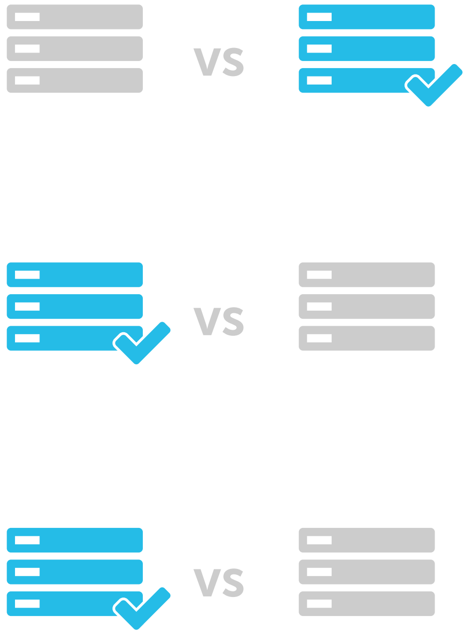
Here are a few questions to ask about benchmarks when you review cloud performance comparisons:

- What benchmarks were used and what do they measure?
- Are the benchmarks reputable and were they run correctly?
- Did the authors interpret the results correctly?
- Are the benchmarks relevant to my workloads?

Compute Instance Selection

Cloud compute services offer multiple types of compute instances and configurations to choose from. To accurately compare services you must first select comparable instance types. (It would be easy to claim that one service is faster than another by simply comparing dissimilar instances. In fact, such apples-to-oranges comparisons often appear in public studies.)

It is often difficult to select compute instances for comparison from multiple providers. Variations in terminology, features, and pricing are many. Often the best you can do when making such selections is to choose instances that are as similar as possible. Factors to consider include number of CPU cores, processor type, amount of memory, extra features and options, and pricing.



Web Server

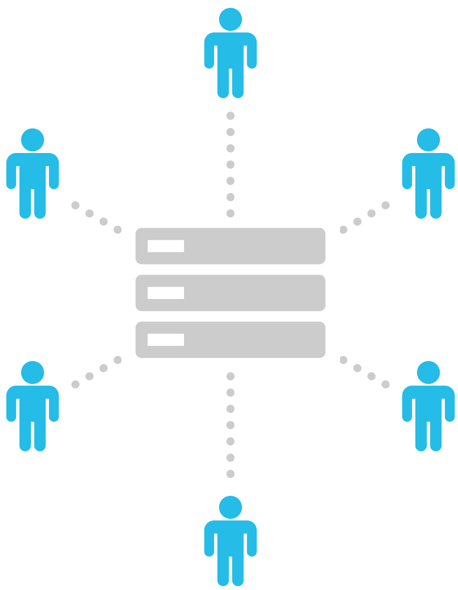
In the comparisons provided in this report, we chose web server instances with fast processing and disk read performance. We also selected local disk storage because web servers don't usually store application data. (Compute services often offer both external and local disk storage options; external disk storage persists beyond the lifespan of a compute instance, while local disk storage does not. Local disk storage is often faster.)

To maximize performance, we selected web server instances from provider services that use current generation CPU hardware and SSD local storage where possible. We also targeted a ratio of 1-2X CPU cores-to-memory to accommodate typical web server memory usage. Because the number of CPU cores directly influenced our processing benchmarks, while amount of memory did not, we matched the number of CPU cores for each provider, while allowing memory to be vary. We chose three web server instances from each provider, based on the following specifications:

Web Server Type	CPU Cores	Memory	Storage
Small	2	2-4 GB	Local
Medium	4	4-8 GB	Local
Large	8	8-16 GB	Local

Database Server

A database server manages persistent application data. Clients interact with a database server to read and write this data - for example to manage a customer database or product catalog. Because this data is critical to the workload, it's best to maintain it on an external disk volume that can be recovered if the compute instance fails. Database server performance is often tied to processing and read + write disk performance. Database servers can also use memory to cache data for improved performance.



We selected database server instances from services that use current generation CPU hardware and external storage where possible. To accommodate typical database server memory usage, we targeted a ratio of 2-4X CPU cores-to-memory. As we did with web server instances, the number of CPU cores was matched for each services and instance type. We chose database server instances from each provider, based on the following specifications:

Database Server Type	CPU Cores	Memory	Storage
Small	4	8-16 GB	50 GB Data Volume
Medium	8	16-32 GB	100 GB Data Volume
Large	16	32-64 GB	200 GB Data Volume

Provider Notes

Some providers offer multiple compute instance classes. This section explains why we chose one type of instance over another for each provider.

Amazon EC2

Amazon Elastic Compute Cloud (EC2) offers a large selection of compute instances, 13 classes in total. Some of these classes have been replaced, but are still offered to support early customers still using them. These previous generation instance classes are not recommended for new users.

Two Amazon EC2 instance classes matched the web and database server criteria listed above:

- M3 class uses current generation Intel Xeon E5-2670 hardware, provides SSD local storage, and uses a ratio of 3-4X CPU cores-to-memory. M3 instances are available with 1-8 CPU cores.
- C3 class uses current generation Intel Xeon E5-2680v2 hardware, provides SSD local storage, and uses a ratio of 2X CPU cores-to-memory. C3 instances are available with 2-32 CPU cores.

In addition to local SSD storage, both Amazon EC2 M3 and C3 instance classes provide the option to use external storage volumes using Amazon Elastic Block Store (Amazon EBS).

The C3 instances best matched the criteria for web and database instances because of its CPU cores-to-memory ratio and availability of 16-core instances for the large database instance type.

Amazon EC2 T2 Instances

On July 1 2014, Amazon announced T2 instances with burstable CPU resources. This class offers 1 or 2 CPU cores with predictable, credit based, CPU bursting. It offers a better value option for bursty workloads (those with periodic peak activity). CPU bursting is nothing new, but the T2 implementation using a predictable bursting model is, and potentially offers good value for workloads that fall within its 10-20% bursting allowance. Because workloads with temporary bursting are common, we have included the t2.medium instance in the small web server CPU performance and value analysis (in addition to the c3.large instance).

DigitalOcean

DigitalOcean offers one class of compute instances. Compute instances use current generation Intel E5 hardware, a 1-2X CPU cores-to-memory ratio, and local SSD storage. DigitalOcean does not support external disks, so a failure of compute instance may result in lost data.

Google Compute Engine

Google Compute Engine uses current generation Intel Sandy Bridge hardware. It is offered in three instance classes, Standard, High Memory, and High CPU.

- Standard instance class uses a ratio of 3-4X CPU cores-to-memory and is available with 1-16 cores.
- High Memory instance class uses a ratio of 6-7X CPU cores-to-memory and is available with 1-16 cores.
- High CPU instance class uses a ratio of slightly less than 1X CPU cores-to-memory and is available with 2-16 cores.

Google Compute Engine does not offer a local SSD disk option. All disks volumes are external to the compute instance.

Because of the CPU cores-to-memory ratio, the High CPU instance class best matches the web server specifications, while the Standard instance class most closely matches the database server.

Rackspace

Rackspace Cloud offers compute instances in two classes, Standard and Performance. The Standard class uses older AMD hardware and generally falls behind in performance. The Performance class uses current generation Intel Xeon E5-2670 hardware. Performance class is available in two flavors:

- Performance 1, which offers a 1X CPU core-to-memory ratio with 1-8 CPU cores.
- Performance 2, which offers a 3-4X CPU core-to-memory ratio with 4-32 CPU cores.

All Performance class compute instances include SSD local storage. Additionally, Rackspace also offers an external storage option called Cloud Block Storage, which is available in either Standard or SSD tiers.

Based of CPU cores-to-memory ratio, Performance 1 was the best match for web server, and Performance 2 for database server instances.

SoftLayer

SoftLayer offers both cloud and traditional dedicated servers. (We excluded dedicated servers from this comparison, because they include set-up fees, extended provisioning times, and monthly commitments and as such generally should not be considered a "cloud service".)

SoftLayer cloud is called CloudLayer. It provides user-selectable CPU cores, memory, and local/external storage options. To provision a new compute instance, users select the number of CPU cores (1-16), amount of memory (1-64 GB), and type of disk volumes (local or external).

Microsoft Azure

Microsoft Azure Virtual Machines currently run on older AMD Opteron 4171 hardware except for compute intensive (A8 and A9) instances which run on Intel E5-2670 processors. Multiple instance classes are currently offered:

- Standard, which has a ratio of less than 2X CPU cores-to-memory and is available with 1-8 CPU cores
- Memory Intensive, which has a ratio of 6-7X CPU cores-to-memory, and is available with 2-8 CPU cores

- CPU Intensive, which has a ratio of 7X CPU cores-to-memory, and is available with 8 or 16 CPU cores

Neither the web server specifications nor the database server specifications call for a 6-7X CPU cores-to-memory ratio, so the Standard instance class is the best fit for both. However, because Standard does not offer a 16 core instance, a CPU Intensive instance was selected.

Microsoft Azure does not offer a local storage option; compute instances use external storage only.

Web Server Comparisons

The following tables list the web server compute instances, configurations, and pricing for each provider. Pricing is hourly and does not reflect commitment-based or prepayment discounts offered by some providers (described later). Pricing for some services varies based on region. The pricing listed below reflects the lowest cost region for each service.

Small Web Server Comparison Matrix

Compute Service	Instance Type	Hourly Price	CPU	Memory	Storage
Amazon EC2	c3.large	$0.105	2 vCPU	3.75 GB	2 x 16 GB SSD
Amazon EC2	t2.medium	$0.068	2 vCPU	4 GB	Off Instance Only
DigitalOcean	4 GB / 2 Cores	$0.06	2 Cores	4 GB	60 GB SSD
Google	n1-highcpu-2	$0.088	2 Cores	1.8 GB	Off Instance Only
Microsoft Azure	Medium (A2)	$0.12	2 Cores	3.5 GB	Off Instance Only
Rackspace	Performance 1 2GB	$0.08	2 vCPU	2 GB	20 GB SSD Data Disk
SoftLayer	2 GB / 2 Cores	$0.083	2 Cores	2 GB	25 GB Local Storage

Medium Web Server Comparison Matrix

Compute Service	Instance Type	Hourly Price	CPU	Memory	Storage
Amazon EC2	c3.xlarge	$0.21	4 vCPU	7.5 GB	2 x 40 GB SSD
DigitalOcean	8 GB / 4 Cores	$0.119	4 Cores	8 GB	80 GB SSD
Google	n1-highcpu-4	$0.176	4 Cores	3.6 GB	Off Instance Only
Microsoft Azure	Large (A3)	$0.24	4 Cores	7 GB	Off Instance Only
Rackspace	Performance 1 4GB	$0.16	4 vCPU	4 GB	40 GB SSD Data Disk
SoftLayer	4 GB / 4 Cores	$0.167	4 Cores	4 GB	25 GB Local Storage

Large Web Server Comparison Matrix

Compute Service	Instance Type	Hourly Price	CPU	Memory	Storage
Amazon EC2	c3.2xlarge	$0.42	8 vCPU	15 GB	2 x 80 GB SSD
DigitalOcean	16 GB / 8 Cores	$0.238	8 Cores	16 GB	160 GB Local SSD
Google	n1-highcpu-8	$0.352	8 Cores	7.2 GB	Off Instance Only
Microsoft Azure	Extra Large (A4)	$0.48	8 Cores	14 GB	Off Instance Only
Rackspace	Performance 1 8GB	$0.32	8 vCPU	8 GB	300 GB Local SSD
SoftLayer	8 GB / 8 Cores	$0.323	8 Cores	8 GB	25 GB Local

Database Server Comparisons

The following tables list the database server compute instances, configurations, and pricing for each provider. Pricing is hourly and does not reflect commitment-based or prepayment discounts offered by some providers (described later). Pricing is for compute instances only and does not include additional cost of external storage. Pricing for some services varies based on region. The pricing listed below reflects the lowest cost region for each service.

Small Database Server Comparison Matrix

Compute Service	Instance Type	Hourly Price	CPU	Memory	Storage
Amazon EC2	c3.xlarge	$0.21	4 vCPU	7.5 GB	50 GB EBS General Purpose (SSD) 50 GB EBS 1500 PIOPS
DigitalOcean	8 GB / 4 Cores	$0.119	4 Cores	8 GB	80 GB Local SSD
Google	n1-standard-4	$0.28	4 Cores	15 GB	50 GB Persistent Disk
Microsoft Azure	Large (A3)	$0.24	4 Cores	7 GB	50 GB LRS Off Instance
Rackspace	Performance 2 15GB	$0.68	4 vCPU	15 GB	100 GB Std Cloud Block Storage 100 GB SSD Cloud Block Storage
SoftLayer	8 GB / 4 Cores	$0.236	4 Cores	8 GB	50 GB SAN

Medium Database Server Comparison Matrix

Compute Service	Instance Type	Hourly Price	CPU	Memory	Storage
Amazon EC2	c3.2xlarge	$0.42	8 vCPU	15 GB	100 GB EBS General Purpose (SSD) 100 GB EBS 3000 PIOPS
DigitalOcean	16 GB / 8 Cores	$0.238	8 Cores	16 GB	160 GB Local SSD
Google	n1-standard-8	$0.56	8 Cores	30 GB	100 GB Persistent Disk
Microsoft Azure	Extra Large (A4)	$0.48	8 Cores	14 GB	100 GB LRS Off Instance
Rackspace	Performance 2 30GB	$1.36	8 vCPU	30 GB	100 GB Std Cloud Block Storage 100 GB SSD Cloud Block Storage
SoftLayer	16 GB / 8 Cores	$0.438	8 Cores	16 GB	100 GB SAN

Large Database Server Comparison Matrix

Compute Service	Instance Type	Hourly Price	CPU	Memory	Storage
Amazon EC2	c3.4xlarge	$0.84	16 vCPU	30 GB	200 GB EBS General Purpose (SSD) 200 GB EBS 4000 PIOPS
DigitalOcean	48 GB / 16 Cores	$0.705	16 Cores	48 GB	480 GB Local SSD
Google	n1-standard-16	$1.12	16 Cores	60 GB	200 GB Persistent Disk
Microsoft Azure	A9	$4.47	16 Cores	112 GB	200 GB LRS Off Instance
Rackspace	Performance 2 60GB	$2.72	16 vCPU	60 GB	200 GB Std Cloud Block Storage 200 GB SSD Cloud Block Storage
SoftLayer	32 GB / 16 Cores	$0.794	16 Cores	32 GB	200 GB SAN

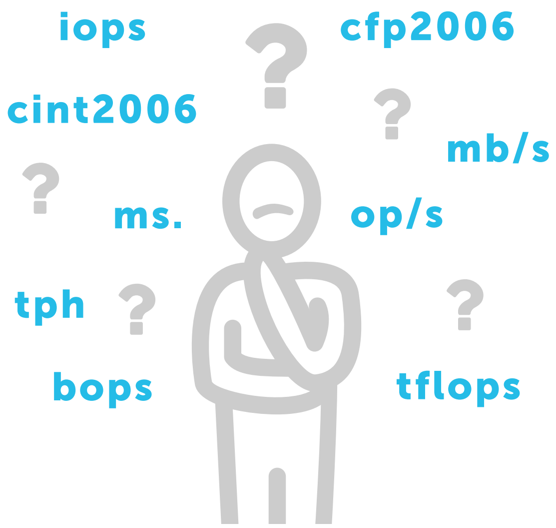
Processor Types

Number of CPU cores and clock speed is not the only determining factor for CPU performance. Processor type also directly influences CPU performance. Newer processors usually support optimizations and faster memory/cache which can improve application performance. Although providers often do not publicly disclose the type of hardware they use (although some do), unless they have configured their virtualization software to hide this information, you can usually look it up in a running compute instance. When running our benchmarks, processor type is one of the attributes we capture. The table below presents the processor types we observed for each service based on the instance selections listed above.

Service	Processor Type	Release Date
Amazon EC2	c3 instances: Intel Xeon E5-2680 v2 2.80GHz	Q3 2013
	t2.medium: Intel Xeon E5-2670 v2 2.50GHz	Q3 2013
DigitalOcean	Not disclosed	
Google	Intel Xeon 2.60GHz (likely E5-2670)	Q1 2012
Microsoft Azure	AMD Opteron 4171 HE	Q2 2010
Azure A8-A9	Intel Xeon E5-2670 2.60 GHz	Q1 2012
Rackspace	Intel Xeon E5-2670 2.60GHz	Q1 2012
SoftLayer	Intel Xeon E3-1270 3.40GHz	Q2 2011
	Intel Xeon E5-2650 2.00GHz	Q1 2012
	Intel Xeon E5-2650 v2 2.60GHz	Q3 2013
	Intel Xeon E5620 2.40GHz	Q1 2010
	Intel Xeon X3470 2.93GHz	Q3 2009
	Intel Xeon X5570 2.93GHz	Q1 2009

Benchmark Selection

After you have picked the compute services and instance types, you must choose the benchmarks you'll use to compare performance. There are hundreds of possible benchmarks to choose from: the challenge is to choose benchmarks that are relevant; to run them correctly; and to understand the metrics they provide. Many published comparisons fail in one or more of these areas. The best benchmarks you can run are based on your actual workloads. If you have the time and resources available, developing your own custom benchmarks is the best method for comparison. If you do not, you should select existing benchmarks that measure performance characteristics that are relevant to your workloads.



Relevant Performance Characteristics

To select relevant benchmarks, you must know what the relevant performance characteristics are to your workloads. These can be narrowed to a few basic types: processing, disk, memory, and network. Performance of your workloads is probably dependent on one or more of these characteristics.

Processing

Processing is a commonly compared performance characteristic. To some extent, all workloads are affected by processing performance. Workloads with intensive math, scientific, or geographic computations are more likely to be affected than those that simply read and write files to disk or transfer data over the network.

Processing performance is usually bound to the number of CPU cores, clock speed, and type of CPU hardware your compute instance runs on. A new CPU model will usually outperform the models it precedes, even with the same number of cores and clock speed.

Disk Read and Write

Disk performance determines how quickly your disk volumes allow applications to interact with data or files. Disk usage patterns often vary by workload. The web server, for example, is likely to perform many reads, but very few writes; while the database server will likely perform both reads and writes. Your workload might also access data using many small pieces in different places, or fewer large pieces in close proximity. The extent to which you should consider disk performance depends on your workload.

Memory

Interacting with data in memory is usually much faster than interacting with data on disk. Because every running application uses memory, it affects performance of all workloads to some extent. For example, to improve performance, a database server may use memory to cache frequently run queries, or to buffer write operations. Newer CPU models usually support faster memory performance due to technological advancements. If your applications have large memory footprints, or use memory for performance optimization, you should factor memory performance more heavily into your comparisons.

Network

Network performance determines how fast your applications can communicate with clients over provider networks. Clients might be users browsing your website over the Internet, or a web server querying data from a database server on the internal network. Network performance is a difficult performance characteristic to compare because there are so many ways to measure it. Generally speaking, you'll want to consider two types of network performance, internal and external.

Internal network performance determines how fast your applications can communicate with other compute instances from the same service, such as between different servers (e.g., web to database servers) or from one service to another (e.g., compute to object storage). When you evaluate internal network performance, consider whether you plan to maintain compute instances in the same or different data centers, other services you'll be using, and the amount and type of network traffic that will occur between then.

External network performance determines how fast your applications can communicate with clients over the Internet. Clients might be end-users or other servers you have in your own data center. Measuring and comparing external network performance is complex because there are many variables to consider, such as where the client is located and the network route a client uses to reach the service. When you evaluate external network performance, consider where the clients are located. This will help to both narrow your testing and comparison criteria, and choose the most optimal service region to use.

Web Server

For our web server workload, we'll compare processing, disk, and external network performance. For disk, we focus on read performance because web servers usually perform primarily read operations. For external network performance, we'll focus on users in North America and Europe using data we collect from the CloudHarmony Speedtest (cloudharmony.com/speedtest). The following table summarizes these selections.

Benchmark	Description	Notes
SPEC CPU 2006	SPEC's industry-standardized, CPU-intensive benchmark suite, stressing a system's processor, memory subsystem and compiler. This benchmark consists of 29 underlying sub-benchmarks that measure CPU integer and floating point performance characteristics.	• SPECint and SPECfp - base/rate tests performed • 1 copy per CPU core • 64-bit binaries • Compiled using Intel compiler suite
fio	fio is used to simulate and measure IO load characteristics based on a wide range of IO load options.	fio workloads included in analysis: • 4k random and sequential read • 16k random and sequential read • 32k random and sequential read • 64k random and sequential read • 128k random and sequential read • 1m random and sequential read Options: • ioengine: libaio • iodepth: optimal for IOPS • rampup time: 5 minutes • runtime: 10 minutes Notes: • Buffers randomized and volume filled prior to testing • Optimal iodepth for each workload calculated during runup • 3 test iterations performed
External Network	A free service we provide that you can use to measure your connectivity to multiple services. Includes latency, downlink, and uplink tests.	Comparisons are based on the latency component of the Speedtest

Database Server

For our database server workload, we'll compare processing, memory, disk, and internal (same data center) network performance. For disk, we focus on read + write performance using a ratio of 80% reads to 20%. Disk performance consistency is also evaluated, because this can be a factor in database performance. The following table summarizes these selections.

Benchmark	Description	Notes
SPEC CPU 2006	SPEC's industry-standardized, CPU-intensive benchmark suite, stressing a system's processor, memory subsystem and compiler. This benchmark consists of 29 underlying sub-benchmarks that measure CPU integer and floating point performance characteristics.	• SPECint and SPECfp - base/rate tests performed • 1 copy per CPU core • 64-bit binaries • Compiled using Intel compiler suite
Stream	This benchmark tests the system memory (RAM) performance.	Tests included: • Add • Copy • Scale • Triad
fio	fio is used to simulate and measure IO load characteristics based on a wide range of IO load options.	fio workloads included in analysis: • 4k random and sequential read/write (80/20) • 16k random and sequential read/write (80/20) • 32k random and sequential read/write (80/20) • 64k random and sequential read/write (80/20) • 128k random and sequential read/write (80/20) • 1m random and sequential read/write (80/20) Options: • ioengine: libaio • iodepth: optimal for IOPS • rampup time: 5 minutes • runtime: 10 minutes Notes: • Buffers randomized and volume filled prior to testing • Optimal iodepth for each workload calculated during runup • 3 test iterations performed
Internal Network	Measures throughput and latency between compute instances from the same server in the same region	

Benchmark Comparisons

In this section we compare performance of web and database compute instances using benchmarks applicable to those workloads. We based all of the analysis on testing performed within the last 30 days prior to publishing this report. We performed multiple test iterations on multiple compute instances of each type and size. We used mean calculations to combine metrics from multiple test executions.

Web Server Benchmark Comparisons

To compare web server performance, we used a combination of CPU, disk read, and external network benchmarks. The proceeding sections present these comparisons.

CPU Performance

We used SPEC CPU 2006 to compare CPU performance. SPEC CPU 2006 is an industry standard benchmark suite used to measure computational performance. It consists of 29 underlying benchmarks, 12 for integer, and 17 for floating-point performance. SPEC CPU 2006 produces separate metrics for both integer and floating-point benchmark sets, and can be run with different configurations. The results below are based on a Base/Rate execution with one copy run for each CPU core. The benchmark suite was compiled using Intel C++ and Fortran compilers (version 12.1). All benchmarks were run within a CentOS 6.x operating system, and used the SSE 4.2 CPU instruction set where supported. Results are listed as "estimates" due to SPEC fair use rules in relation to reporting.

CPU Performance Variability

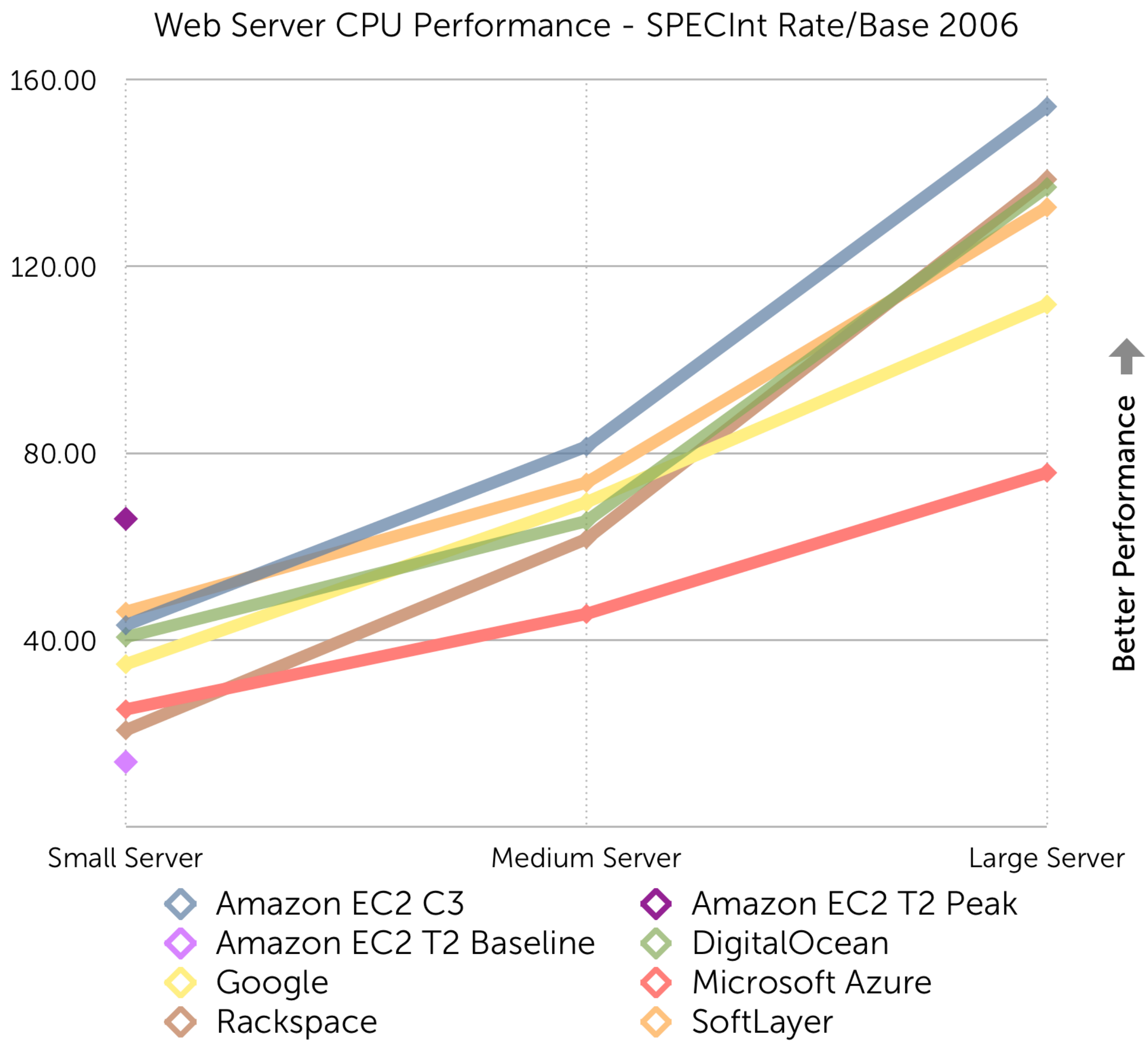
CPU performance variability came in two forms. The first was variability between benchmark runs on the same compute instance. This may be an indication of oversubscription or use of floating CPU core/thread assignments. The second was variability from different compute instances of the same type. This is often caused when a provider uses multiple hardware types (as was the case for SoftLayer). CPU performance variability was apparent by irregular SPEC CPU 2006 result metrics and is presented in the analysis below using the standard deviation of those metrics. In order to compensate for such variability, the SPEC CPU 2006 metrics presented below are based on the mean minus standard deviation from all runs for each instance type.

CPU Performance Comments

For the Amazon EC2 small web server comparisons we included results for both c3.large and t2.medium instances. Because t2.medium is a burstable instance, both peak (burst) and baseline metrics are included. The t2.medium peak metrics are based on tests run entirely in burst mode, while baseline metrics are based on tests run entirely out of burst mode. We accomplished this by allowing instances used for burst testing to maximize CPU credits prior to testing, and by draining CPU credits on instances prior to baseline testing.

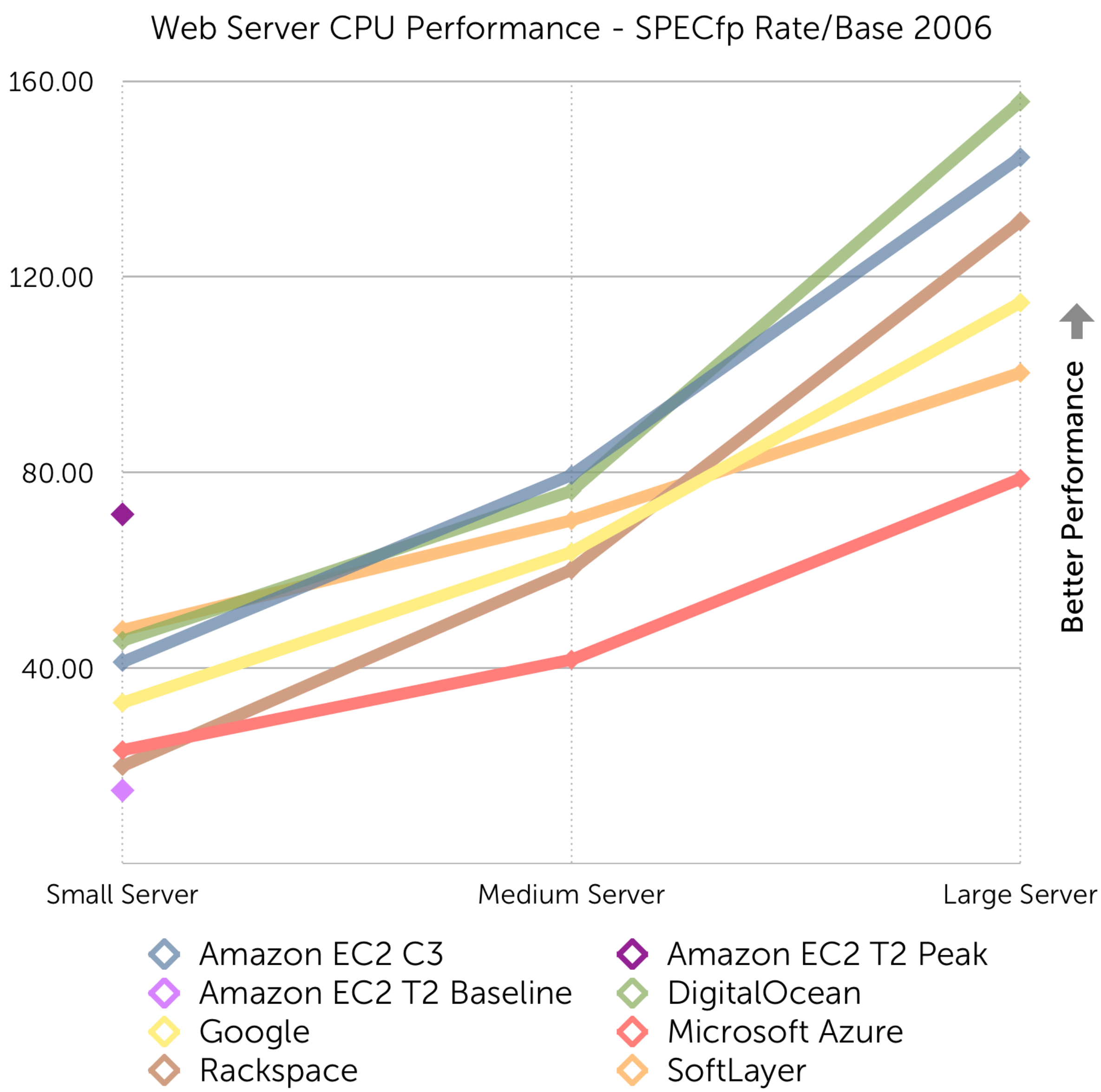
Rackspace compute instances demonstrated high variability across runs on the same instance. This may be due to use of floating CPU core/thread assignment. SoftLayer results were irregular due to use of different CPU models (6 models observed during testing). The best performing SoftLayer instances were those running on Intel Ivy Bridge (E5-2650 v2) CPU models. Microsoft Azure CPU performance lagged behind the others because they use older AMD 4171 processors.

SPECint Rate/Base 2006 (estimate)



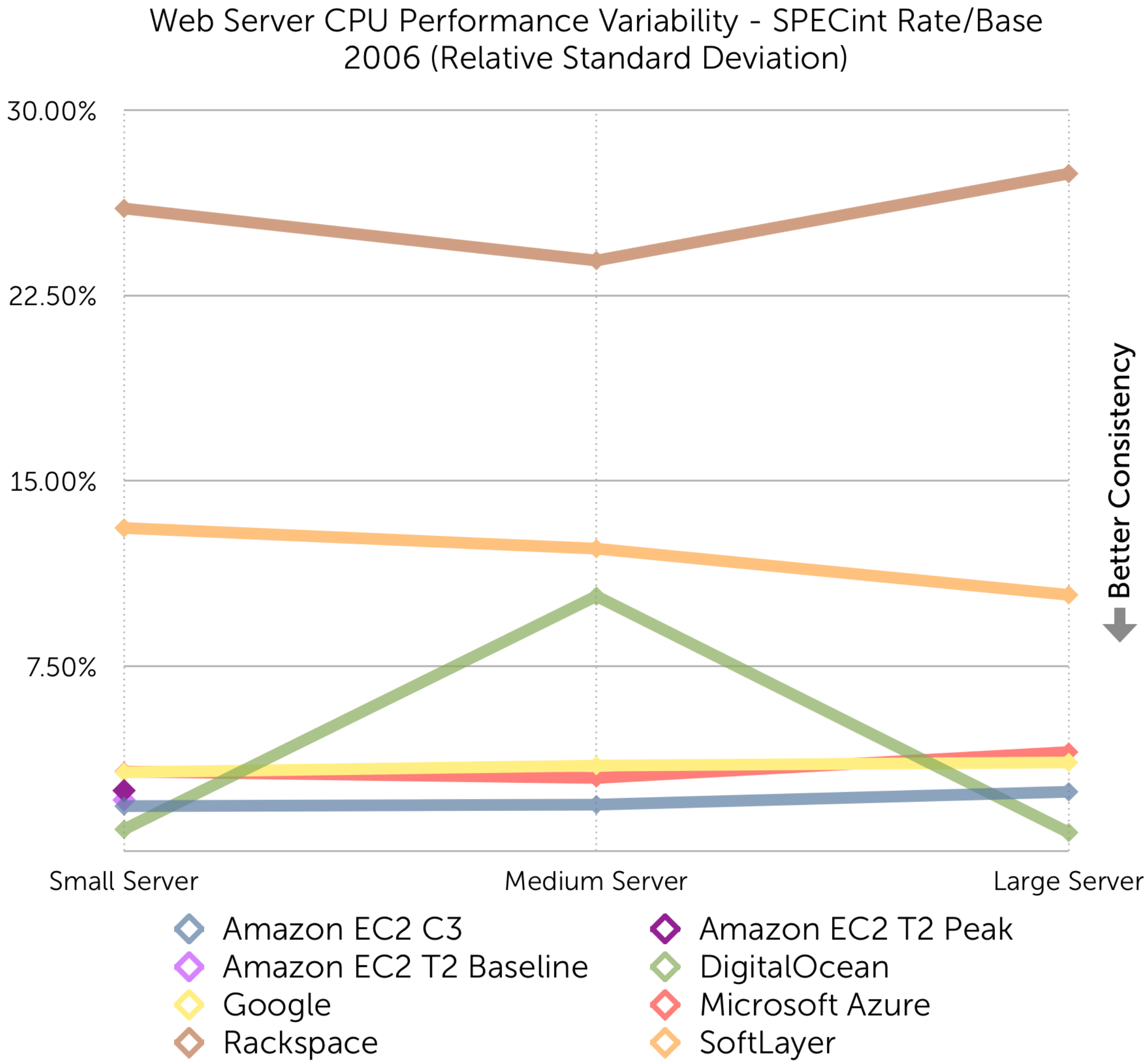
Server Type	Amazon EC2 C3	Amazon EC2 T2 Peak	Amazon EC2 T2 Baseline	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Server	43.19	65.96	13.92	40.64	34.85	25.16	20.71	46.06
Medium Server	81.43	NA	NA	65.46	69.50	45.60	61.63	73.71
Large Server	154.19	NA	NA	136.95	111.82	75.82	138.61	132.64

SPECfp Rate/Base 2006 (estimate)



Server Type	Amazon EC2 C3	Amazon EC2 T2 Peak	Amazon EC2 T2 Baseline	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Server	41.23	71.44	15.00	45.59	32.91	23.22	19.97	47.80
Medium Server	79.47	NA	NA	76.22	63.71	41.72	60.11	70.20
Large Server	144.43	NA	NA	155.81	114.72	78.70	131.35	100.37

CPU Performance Variability



Server Type	Amazon EC2 C3	Amazon EC2 T2 Peak	Amazon EC2 T2 Baseline	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Server	1.83%	2.46%	2.08%	0.89%	3.20%	3.23%	26.02%	13.09%
Medium Server	1.89%	NA	NA	10.33%	3.47%	2.97%	23.91%	12.25%
Large Server	2.41%	NA	NA	0.76%	3.60%	4.02%	27.43%	10.38%

Disk Performance

We used the *fio* benchmark to measure disk performance. In total, we ran three test iterations each, consisting of 36 workloads on each compute instance. The total testing duration was about 40 hours per compute instance. The focus of the web server results provided below is two sets of 100-percent read workload, sequential and random. Testing was optimized for fastest IO using the asynchronous Linux *libaio* engine. Asynchronous IO allows applications to submit operations without waiting for a response. It is usually faster than synchronous IO because applications can submit non-blocking requests to an IO queue, and the operating system will work off that queue in the fastest and most optimal way possible.

When using fio with asynchronous IO, one configurable parameter is the maximum number of requests to place in the queue before waiting for some to be processed. To maximize IO test results, it is best to set this parameter to a value that keeps enough requests in the queue so there are always some pending, but not so many that they overwhelm the system with queuing requests which can reduce IO performance.

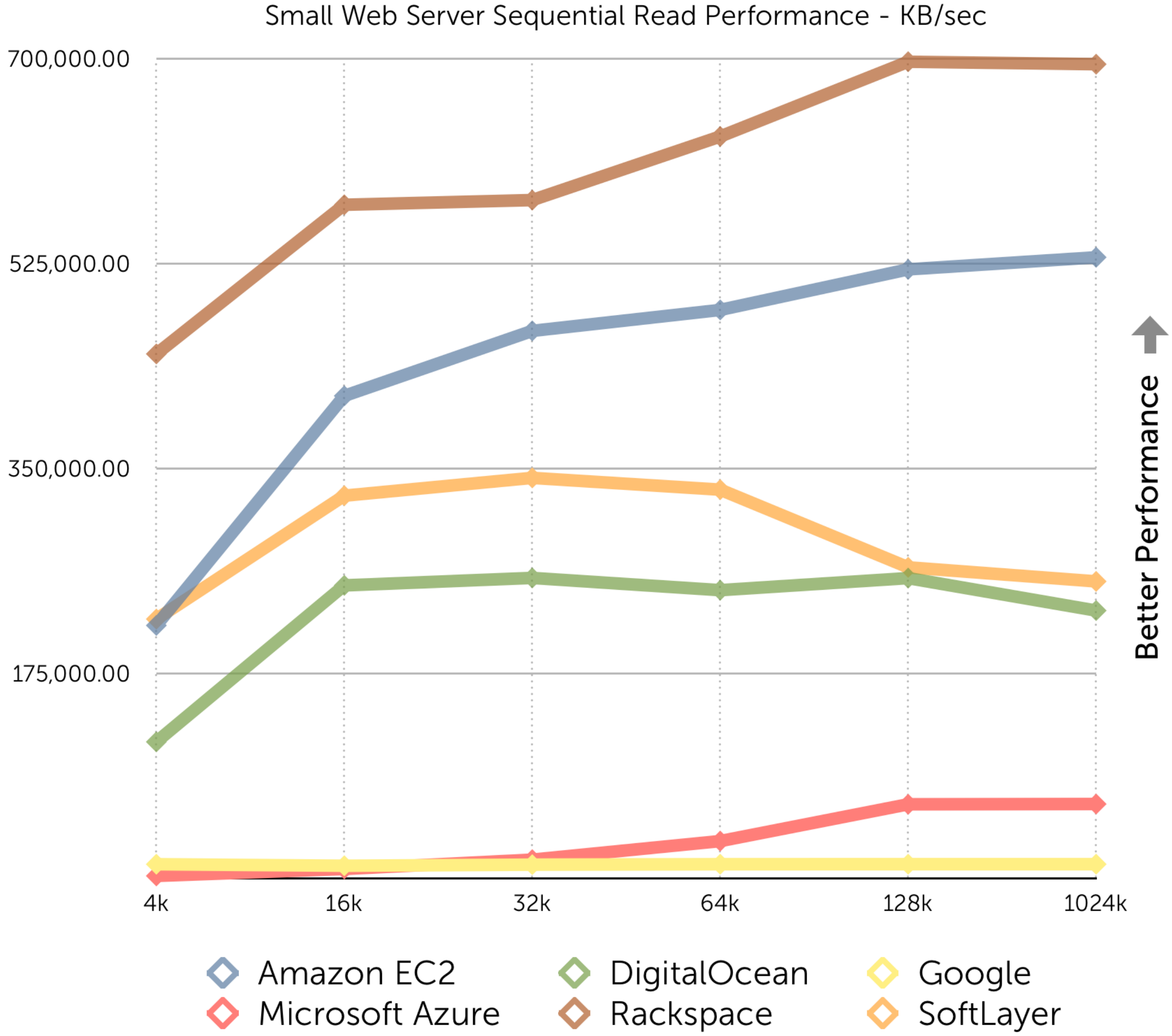
To optimize for IO in our testing, we ran some pre-tests to determine the most optimal *iodepth* for each of the 36 workloads. We ran sets of three tests using incrementing *iodepth* settings (1, 2, 4, 8, etc.) until IO doesn't increase by at least 3 percent. We use these *iodepth* settings during the testing iterations to produce maximized IO for each service.

The analysis below presents IO bandwidth in kilobytes per second for sequential and random read workloads. A higher value represents faster disk performance. We also present IO consistency metrics. These metrics are based on the relative standard deviation of IO during testing, represented as a percentage. A higher value in this analysis represents less consistent IO performance, which can negatively affect application performance.

Disk Performance Comments

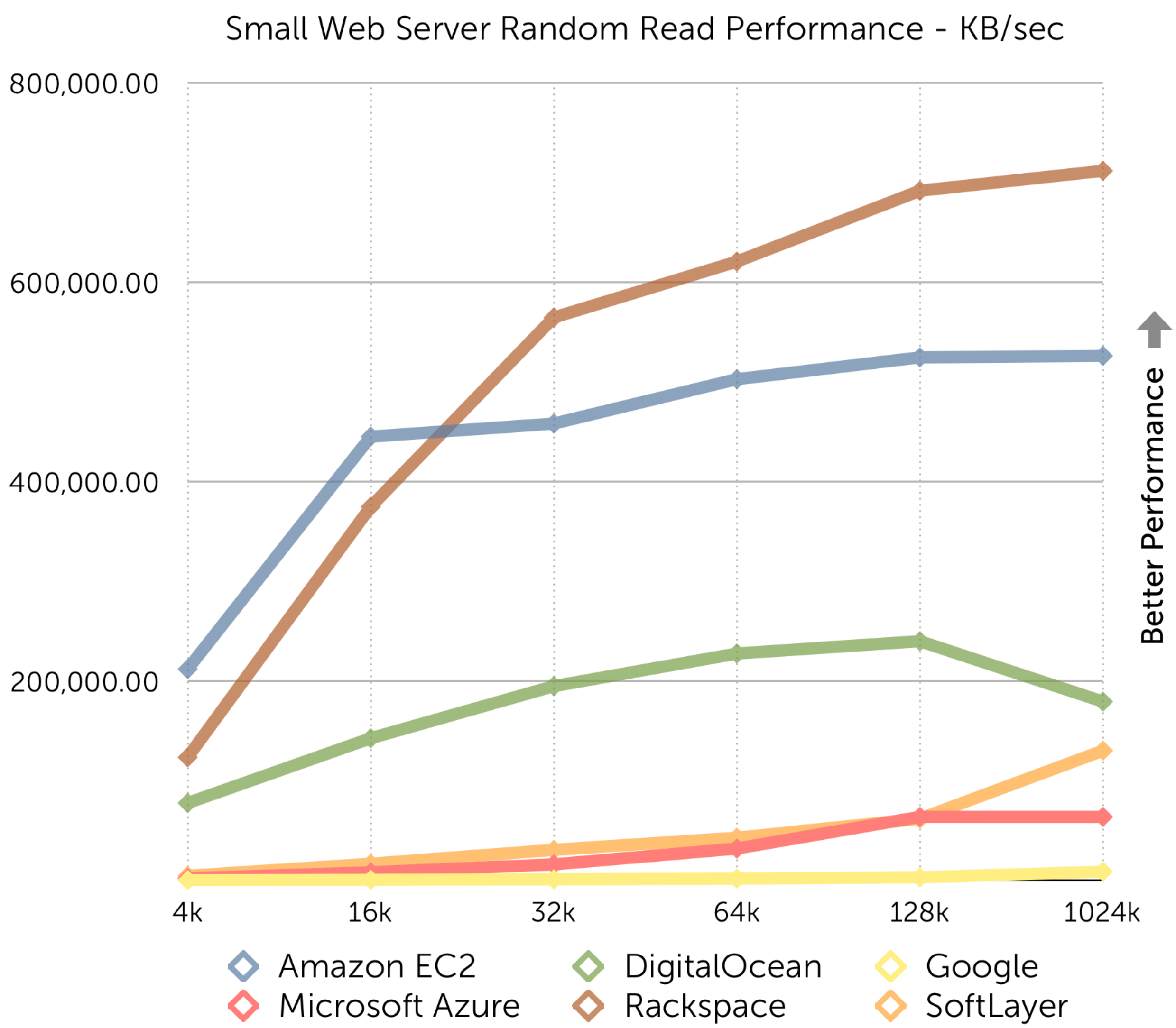
Rackspace and Amazon EC2 SSD performance was generally fastest and most consistent. SoftLayer performance and consistency was also very good, particularly considering that SoftLayer is not an SSD platform. SoftLayer might be using SSD cache to improve performance. DigitalOcean IO performance mostly lagged behind Rackspace and Amazon EC2 in all tests, despite being a pure SSD platform. DigitalOcean IO was also the least consistent, demonstrating high variability during most tests. Neither Google nor Microsoft Azure offer local storage or SSD, so their IO performance predictably lagged behind.

Sequential Read Disk Performance - Small Web Server



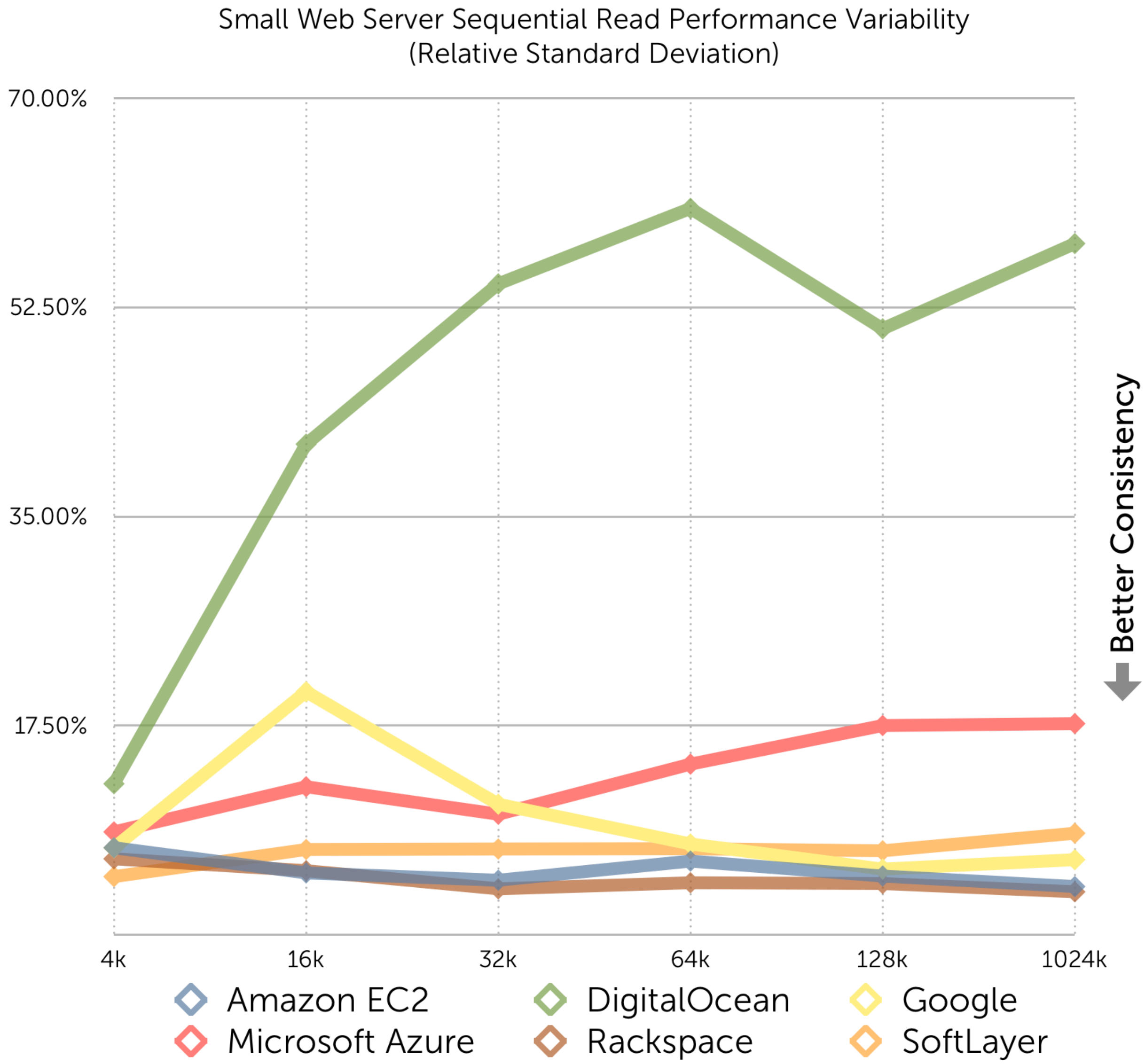
Block Size	Amazon EC2 (KB/s)	DigitalOcean (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace (KB/s)	SoftLayer (KB/s)
4k	215,981.00	116,522.50	11,949.67	1,987.33	447,985.00	221,440.00
16k	412,133.00	250,202.50	10,651.33	7,923.00	575,406.67	326,851.33
32k	467,652.00	256,596.50	11,609.33	15,864.33	579,412.33	342,076.33
64k	485,581.00	246,126.00	11,973.00	31,734.00	633,761.33	332,007.67
128k	520,219.00	256,441.50	11,999.00	63,172.67	697,647.67	265,651.67
1024k	530,532.00	228,696.50	11,999.33	63,411.33	695,529.33	253,619.00

Random Read Disk Performance - Small Web Server



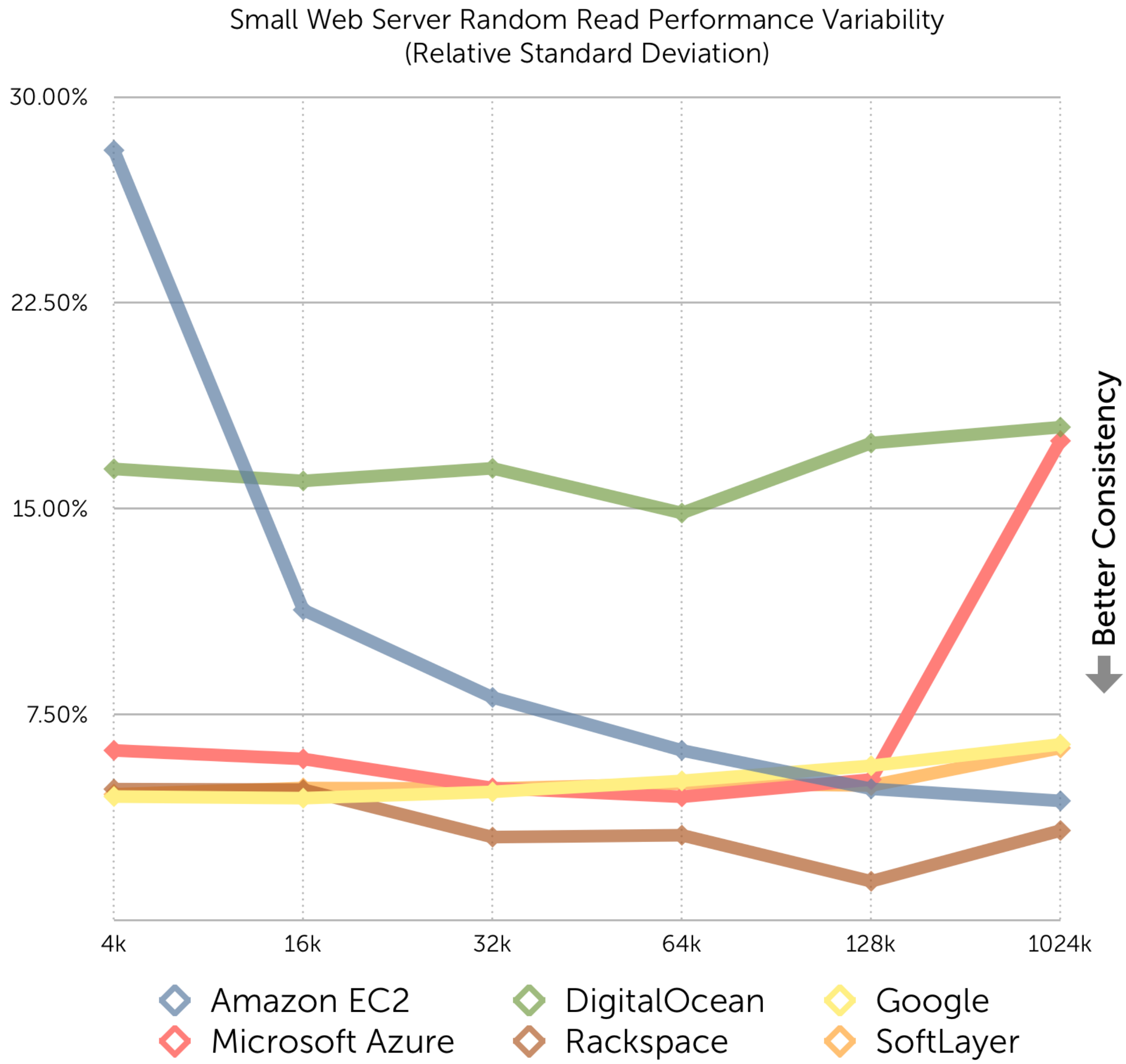
Block Size	Amazon EC2 (KB/s)	DigitalOcean (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace (KB/s)	SoftLayer (KB/s)
4k	211,874.00	77,492.00	117.00	1,986.00	123,370.00	4,409.33
16k	445,217.00	142,487.50	455.00	7,947.67	374,846.33	16,533.67
32k	458,249.00	194,739.00	877.00	15,902.00	564,718.00	30,498.33
64k	502,847.00	227,502.50	1,636.33	31,798.67	620,887.33	42,610.33
128k	524,680.00	239,898.50	2,881.33	63,594.67	691,939.67	61,722.00
1024k	526,222.00	179,227.50	8,600.33	63,547.00	711,957.67	130,037.00

Sequential Disk Read Performance Consistency - Small Web Server



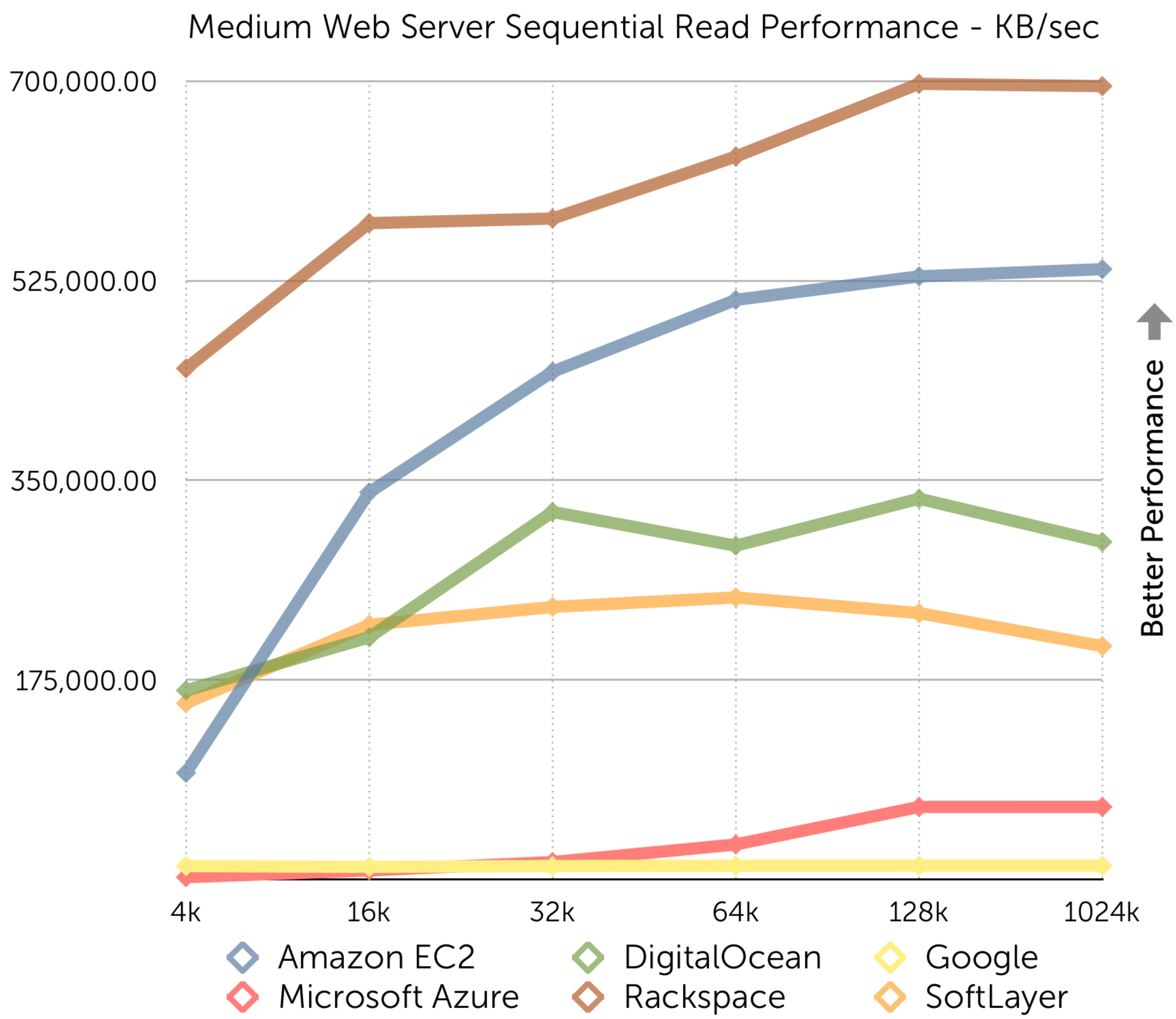
Block Size	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
4k	7.29%	12.64%	7.28%	8.61%	6.36%	4.87%
16k	5.18%	41.05%	20.34%	12.39%	5.37%	7.14%
32k	4.57%	54.46%	10.93%	10.09%	3.85%	7.18%
64k	6.17%	60.77%	7.60%	14.28%	4.36%	7.21%
128k	4.92%	50.70%	5.50%	17.51%	4.29%	7.03%
1024k	4.04%	57.83%	6.29%	17.67%	3.60%	8.50%

Random Disk Read Performance Consistency - Small Web Server



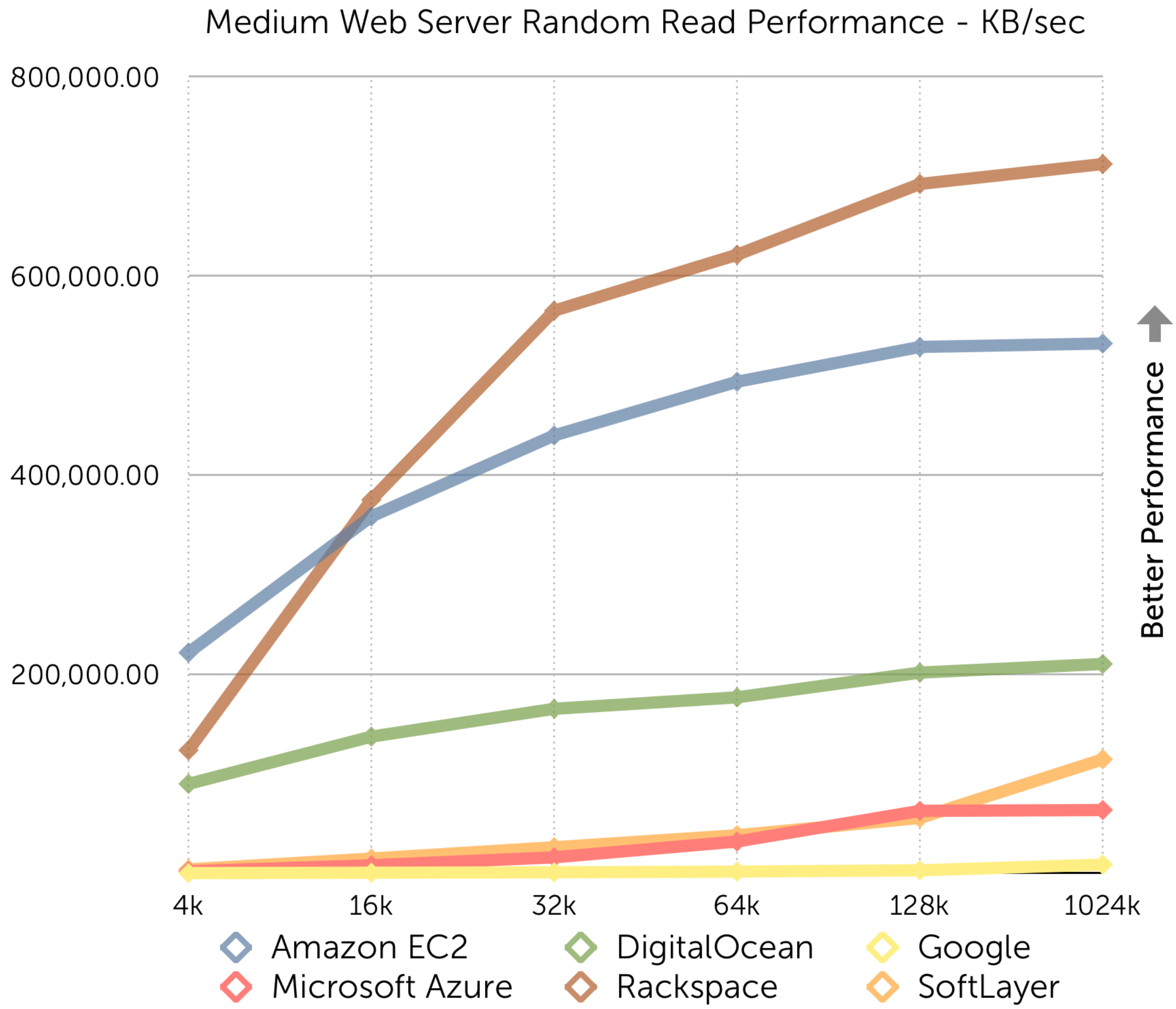
Block Size	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
4k	28.06%	16.45%	4.51%	6.19%	4.78%	4.59%
16k	11.31%	16.01%	4.45%	5.88%	4.76%	4.82%
32k	8.12%	16.47%	4.68%	4.82%	3.03%	4.80%
64k	6.19%	14.84%	5.08%	4.50%	3.10%	4.99%
128k	4.79%	17.39%	5.63%	5.13%	1.42%	4.91%
1024k	4.35%	17.97%	6.41%	17.47%	3.27%	6.28%

Sequential Read Disk Performance - Medium Web Server



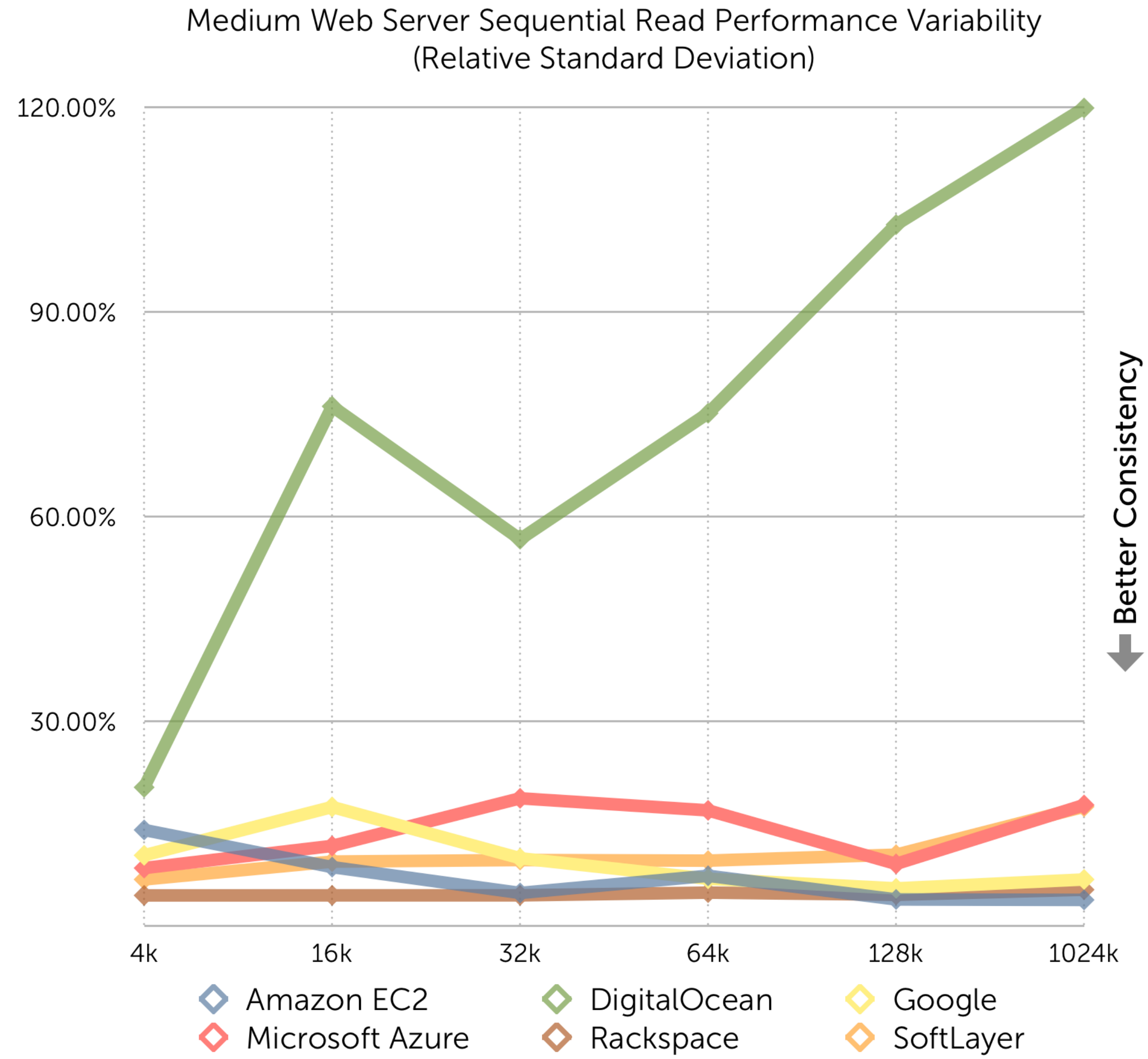
Block Size	Amazon EC2 (KB/s)	DigitalOcean (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace (KB/s)	SoftLayer (KB/s)
4k	93,396.00	165,736.50	11,729.33	1,986.67	447,985.00	154,583.67
16k	339,353.00	212,246.00	10,995.00	7,881.67	575,406.67	223,191.00
32k	445,131.00	322,232.00	11,732.67	15,205.00	579,412.33	238,901.00
64k	508,136.00	292,599.50	11,982.67	30,649.00	633,761.33	247,266.67
128k	528,747.00	333,835.00	11,999.00	63,534.33	697,647.67	233,523.33
1024k	535,087.00	295,859.50	11,999.67	63,547.33	695,529.33	204,570.33

Random Read Disk Performance - Medium Web Server



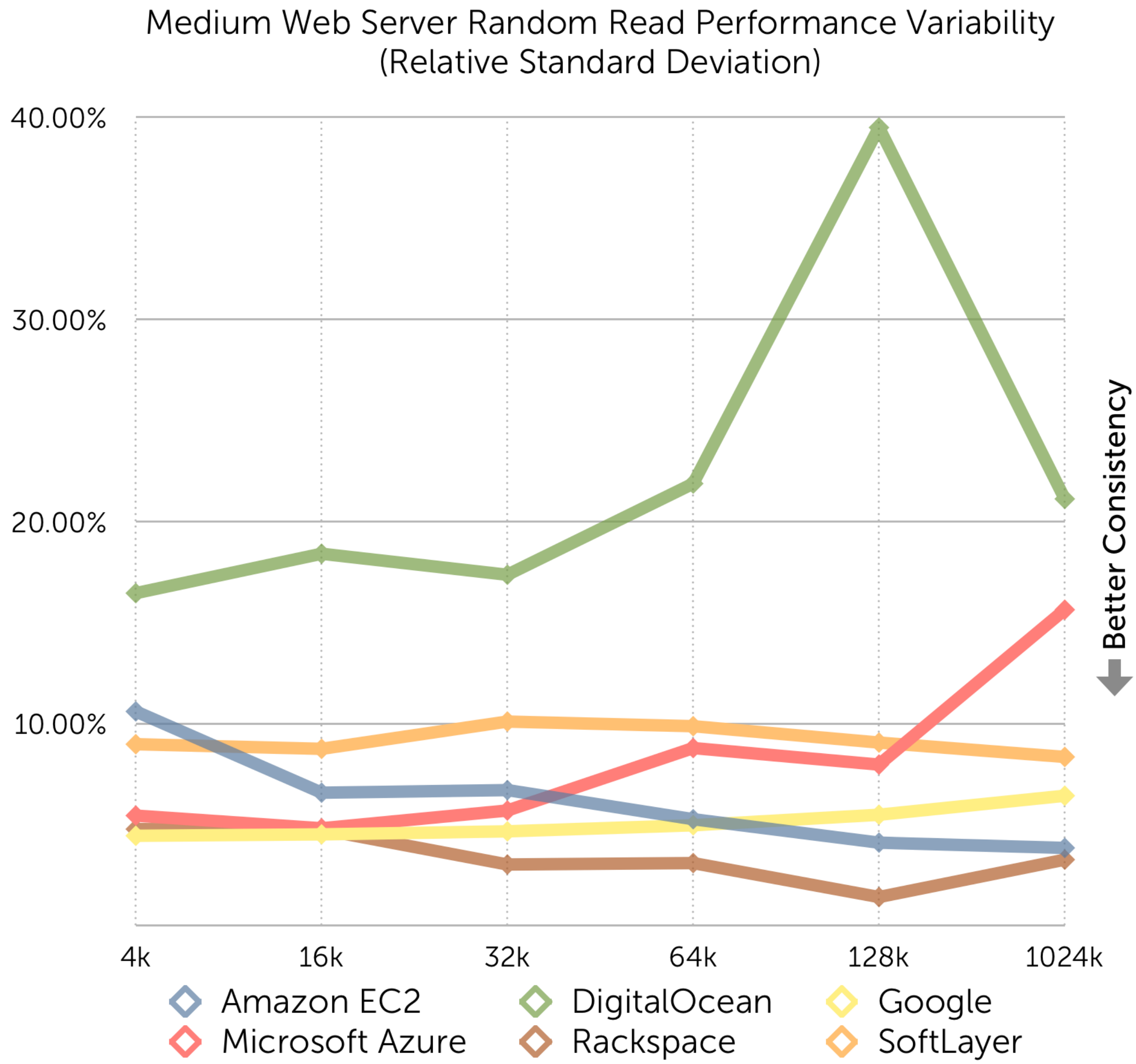
Block Size	Amazon EC2 (KB/s)	DigitalOcean (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace (KB/s)	SoftLayer (KB/s)
4k	221,405.00	89,675.50	117.00	1,986.67	123,370.00	3,836.00
16k	358,128.00	137,088.50	455.00	7,946.67	374,846.33	14,711.00
32k	439,500.00	165,039.50	877.67	15,896.33	564,718.00	26,034.00
64k	493,392.00	176,538.50	1,635.67	31,791.33	620,887.33	38,067.67
128k	528,302.00	201,373.50	2,881.67	62,581.33	691,939.67	55,443.00
1024k	531,785.00	210,048.00	8,603.00	63,426.67	711,957.67	114,374.00

Sequential Disk Read Performance Consistency - Medium Web Server



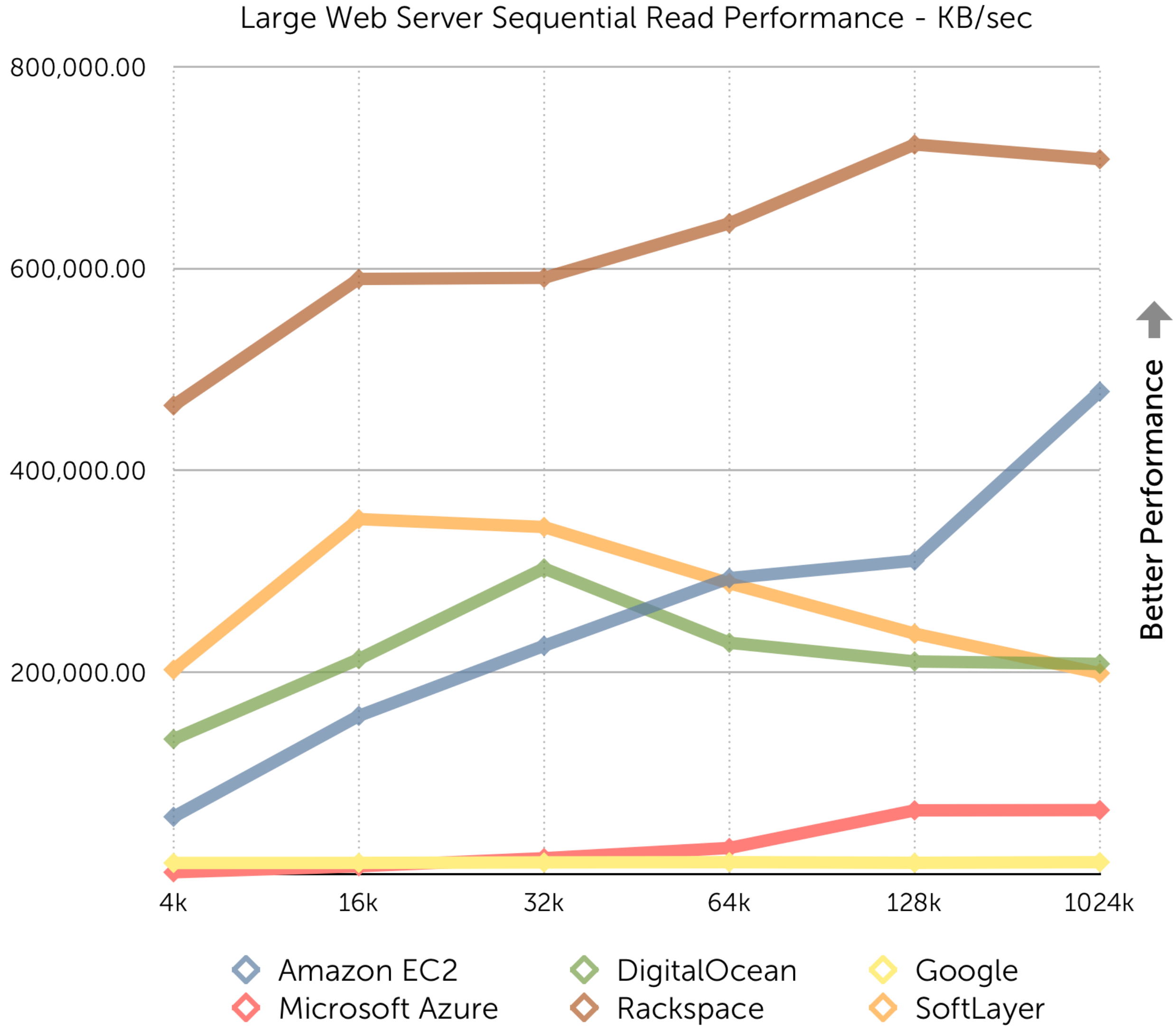
Block Size	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
4k	14.03%	20.28%	10.31%	8.44%	4.43%	6.77%
16k	8.58%	76.15%	17.40%	11.69%	4.44%	9.39%
32k	4.79%	56.67%	9.88%	18.67%	4.44%	9.62%
64k	7.31%	75.14%	6.87%	16.90%	4.84%	9.53%
128k	3.84%	102.80%	5.46%	9.05%	4.52%	10.40%
1024k	3.77%	119.90%	6.77%	17.69%	5.27%	17.44%

Random Disk Read Performance Consistency - Medium Web Server



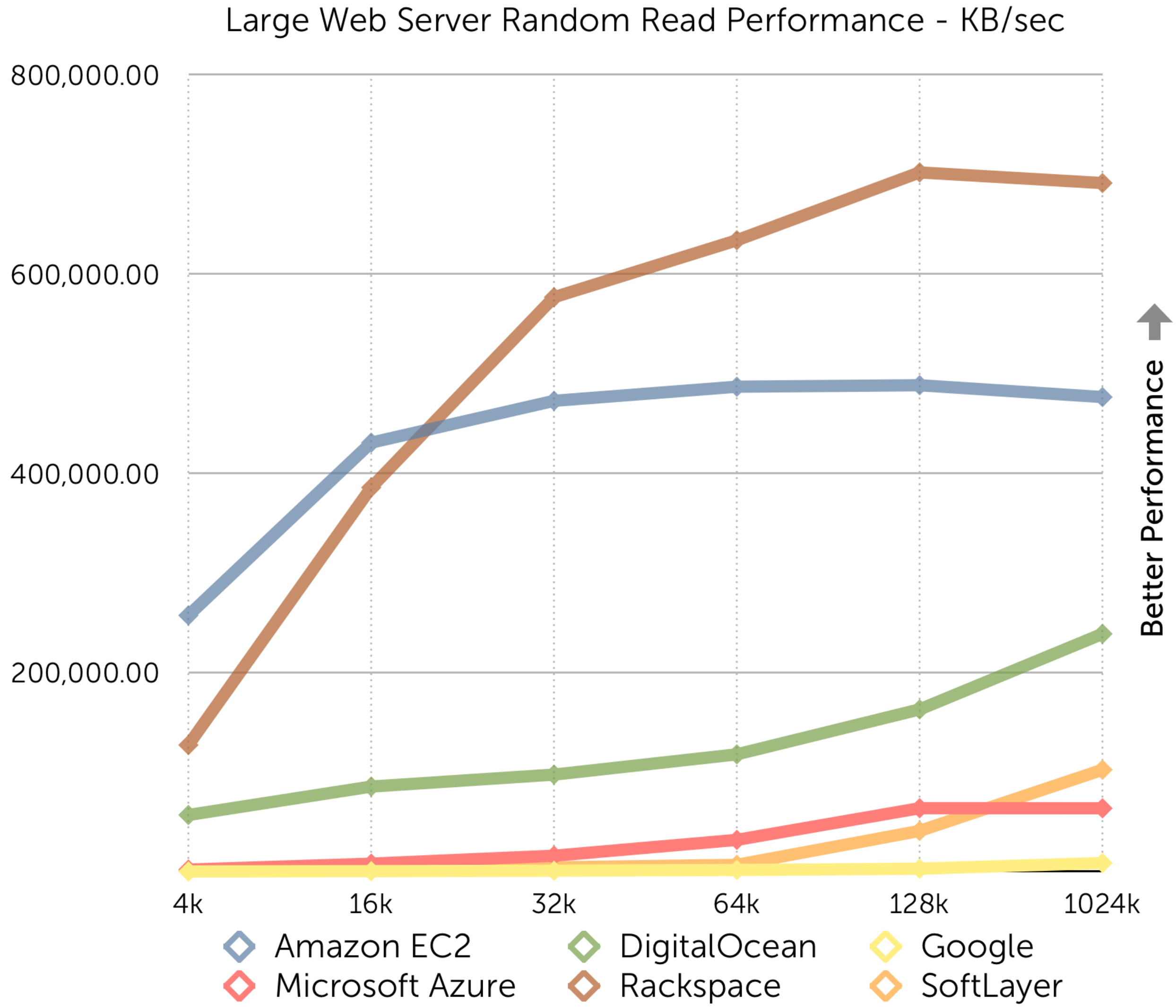
Block Size	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
4k	10.60%	16.45%	4.45%	5.45%	4.78%	8.98%
16k	6.58%	18.40%	4.52%	4.82%	4.76%	8.75%
32k	6.71%	17.36%	4.67%	5.70%	3.03%	10.10%
64k	5.27%	21.87%	4.97%	8.79%	3.10%	9.87%
128k	4.11%	39.48%	5.49%	7.97%	1.42%	9.06%
1024k	3.85%	21.11%	6.43%	15.63%	3.27%	8.35%

Sequential Read Disk Performance - Large Web Server



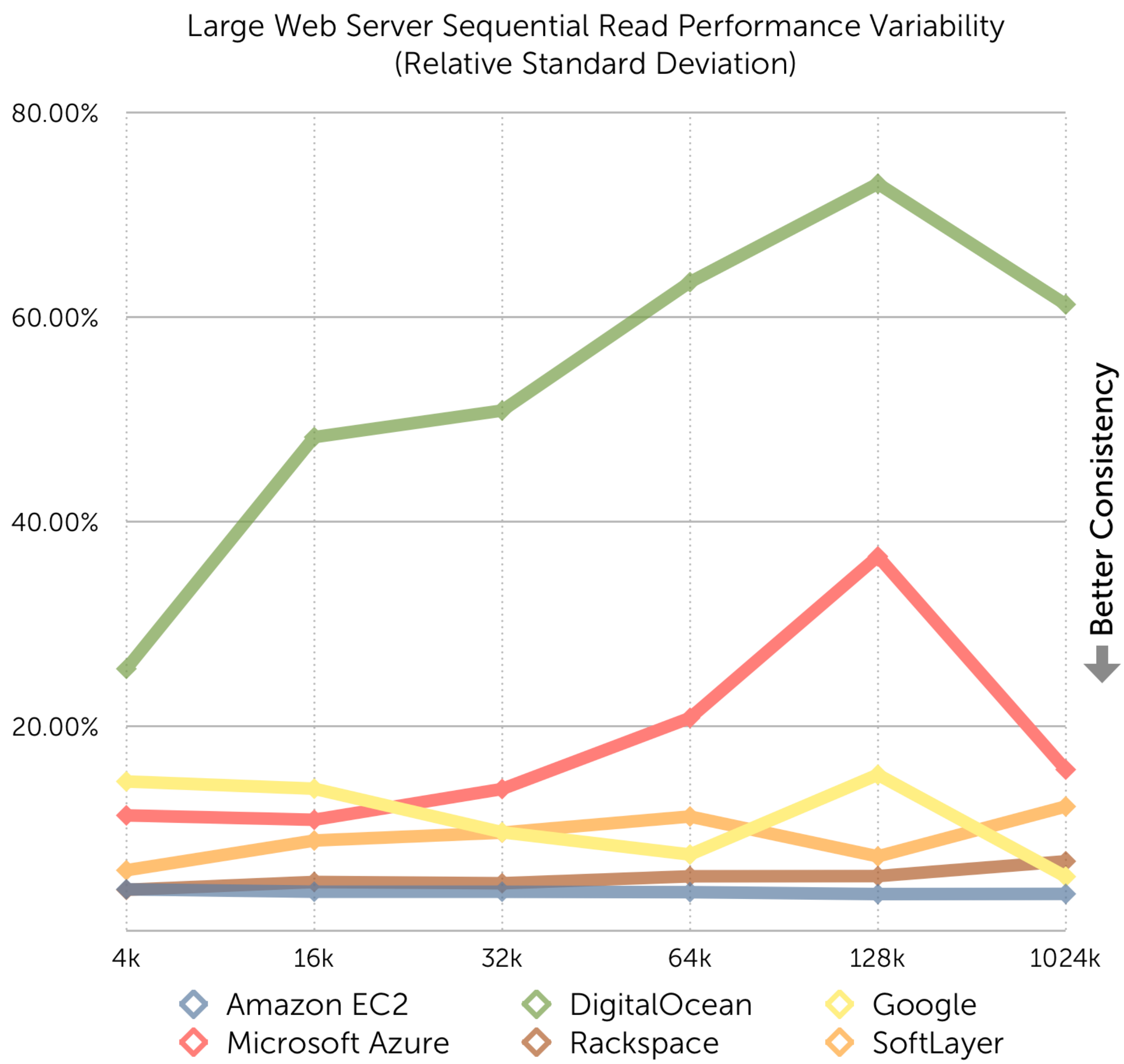
Block Size	Amazon EC2 (KB/s)	DigitalOcean (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace (KB/s)	SoftLayer (KB/s)
4k	56,914.00	133,872.00	11,195.00	1,988.00	464,369.67	202,642.00
16k	156,913.67	213,119.00	11,269.67	7,944.67	589,718.33	351,802.33
32k	226,340.33	302,935.50	11,739.67	15,909.33	590,765.00	343,951.00
64k	293,307.33	229,310.00	11,989.33	26,051.00	644,737.00	288,257.33
128k	310,649.67	210,818.50	11,347.33	63,287.67	722,859.67	238,393.00
1024k	478,129.33	208,381.50	11,999.33	63,594.00	708,262.33	199,117.00

Random Read Disk Performance - Large Web Server



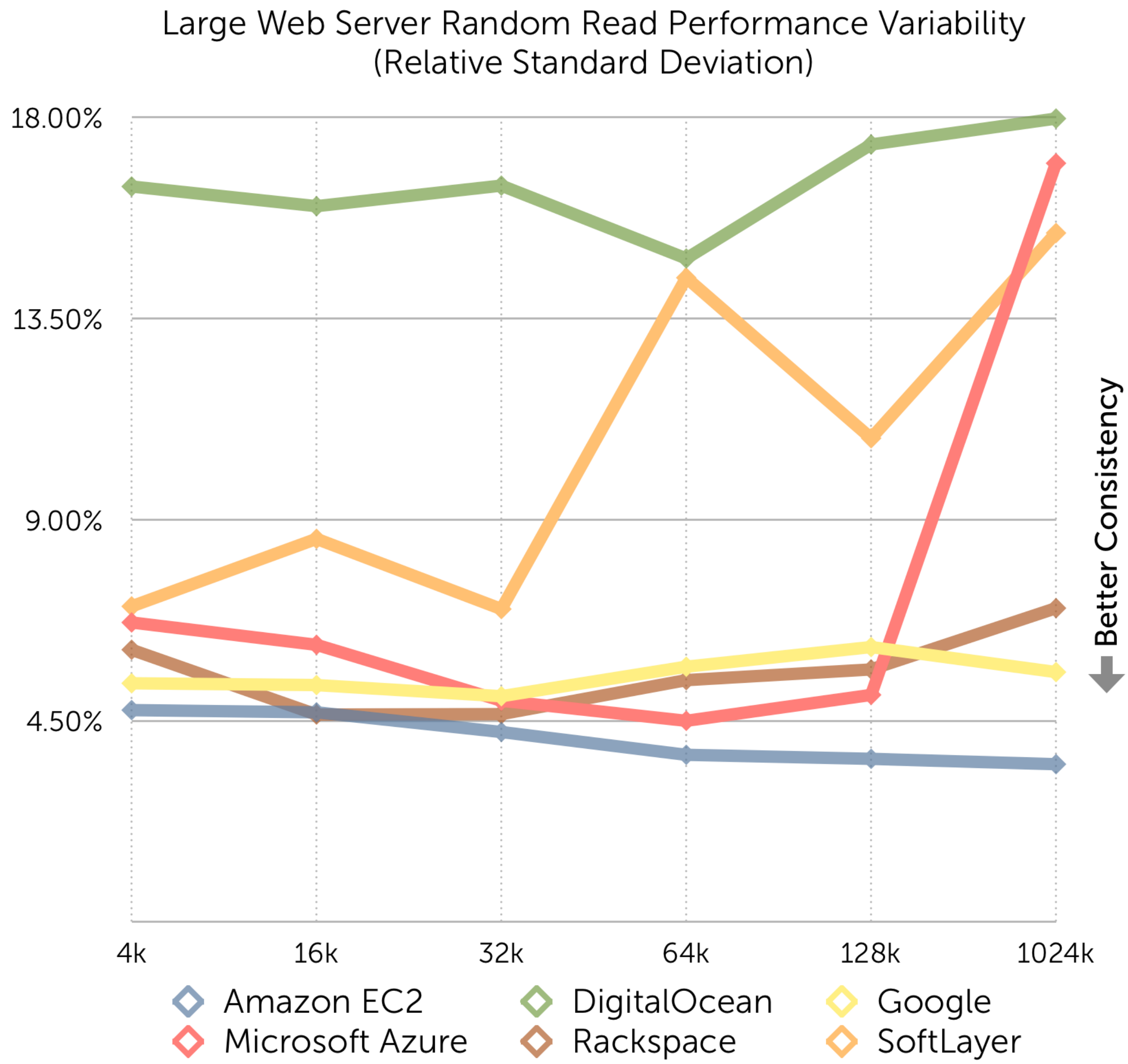
Block Size	Amazon EC2 (KB/s)	DigitalOcean (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace (KB/s)	SoftLayer (KB/s)
4k	257,071.33	56,783.00	117.00	1,988.33	126,999.00	564.33
16k	430,444.67	85,320.50	455.00	7,953.33	385,387.33	2,008.33
32k	472,418.67	97,311.50	877.33	15,910.00	576,393.00	4,235.33
64k	486,594.00	117,648.00	1,637.00	31,811.67	633,572.00	6,780.33
128k	488,087.67	162,757.50	2,880.67	63,602.67	701,665.00	40,815.67
1024k	476,224.33	238,599.50	8,596.67	63,502.67	690,980.00	102,203.67

Sequential Disk Read Performance Consistency - Large Web Server



Block Size	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
4k	4.03%	25.60%	14.58%	11.26%	3.98%	5.89%
16k	3.79%	48.26%	13.87%	10.83%	4.75%	8.80%
32k	3.79%	50.86%	9.61%	13.84%	4.62%	9.57%
64k	3.75%	63.42%	7.43%	20.79%	5.30%	11.15%
128k	3.56%	73.05%	15.24%	36.59%	5.32%	7.25%
1024k	3.58%	61.24%	5.27%	15.74%	6.77%	12.13%

Random Disk Read Performance Consistency - Large Web Server



Block Size	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
4k	4.74%	16.45%	5.34%	6.70%	6.09%	7.06%
16k	4.69%	16.01%	5.30%	6.20%	4.64%	8.57%
32k	4.25%	16.47%	5.05%	4.95%	4.65%	7.00%
64k	3.74%	14.84%	5.71%	4.50%	5.41%	14.41%
128k	3.65%	17.39%	6.15%	5.08%	5.65%	10.82%
1024k	3.53%	17.97%	5.59%	16.97%	7.02%	15.41%

External Network Performance

The data below is based on data collected and summarized from the CloudHarmony speedtest (http://cloudharmony.com/speedtest), which is a public test that you can run in your browser to test network connectivity to cloud services. Thousands of users run this test each month. The analysis in this section summarizes data collected from the latency test between January and May 2014.

Service Regions Used

When you run the speedtest, you can select any service region for testing. Usually, the region with the best network performance is the one closest to the user. We collect data for all service regions, but for the purpose of this analysis, we picked the best performing region for each service and continent. The table below lists the region selections we made for each service.

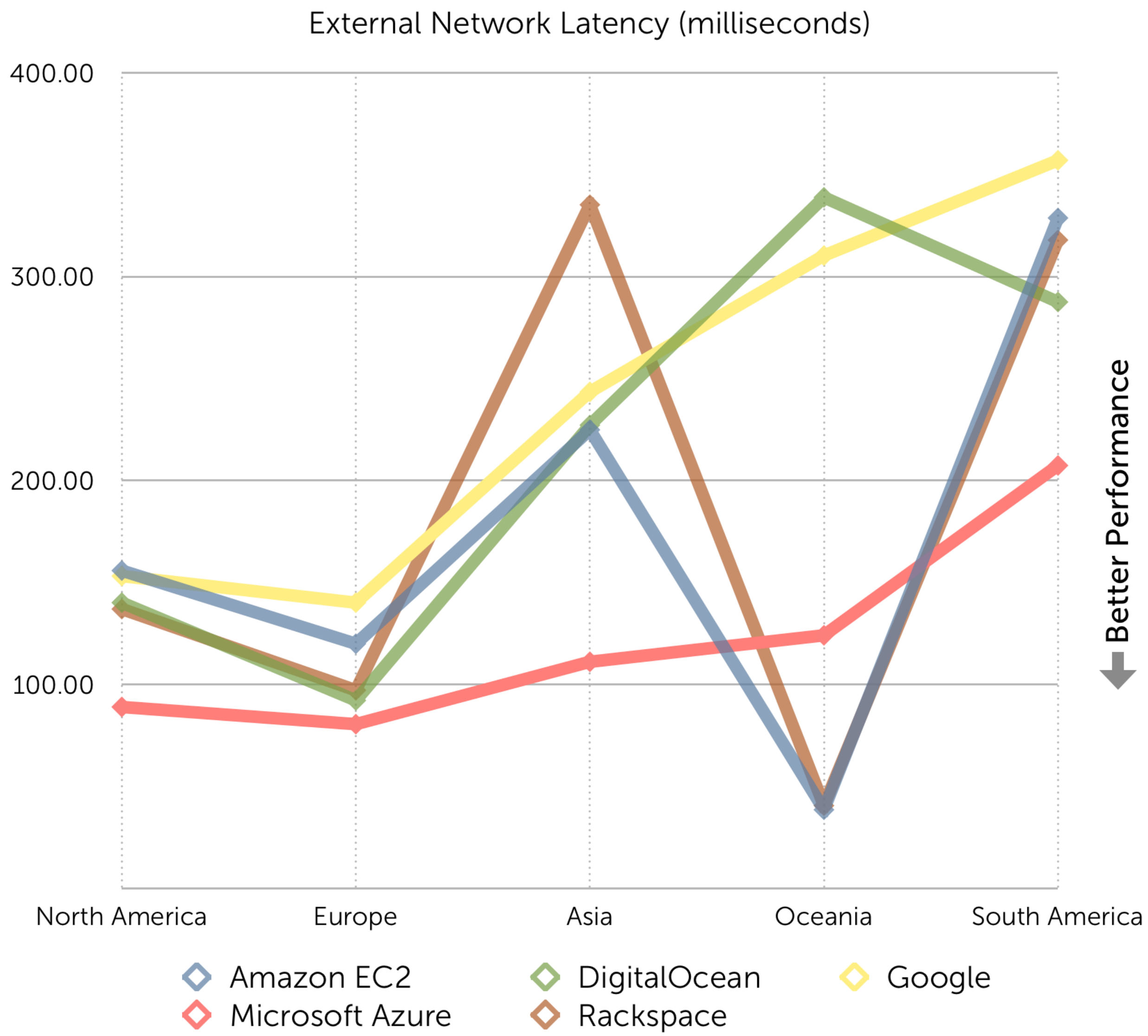
Region	Amazon EC2	DigitalOcean	Google	Microsoft	Rackspace
North America	us-east-1	NY2	us-central1	South Central US	ORD
Europe	eu-west-1	AMS1	europe-west1	West Europe	LON
Asia	ap-southeast-1	SG1	asia-east1	South East Asia	HKG
Oceania	ap-southeast-2	SFO1	asia-east1	South East Asia	SYD
South America	sa-east-1	NY2	us-central1	South Central US	DFW

External Network Performance Comments

Performance was similar in North America and Europe. Providers distinguished themselves most in Asia and Oceania. Because Google and DigitalOcean do not have an Oceania data center, performance was worse in this continent. Performance was equally poor for all services in South America where network connectivity is less reliable.

External Network Latency

The following data is based on tests from thousands of unique users in each continent. SoftLayer is excluded, because we don't have a testing setup currently.



Region	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace
North America	155.86	140.15	153.18	89.01	137.01
Europe	119.99	92.22	140.13	80.58	97.22
Asia	224.88	227.17	243.47	111.15	335.20
Oceania	38.62	338.84	310.30	124.13	40.64
South America	328.72	287.54	357.00	207.40	317.90

Database Server Benchmark Comparisons

To compare database server performance, we used a combination of CPU, memory disk read/write, and internal network benchmarks. The proceeding sections present these comparisons.

CPU Performance

As we did for web server instances, we used SPEC CPU 2006 to analyze and compare CPU performance on database server compute instances. Results are listed as “estimates” due to SPEC fair use rules in relation to reporting.

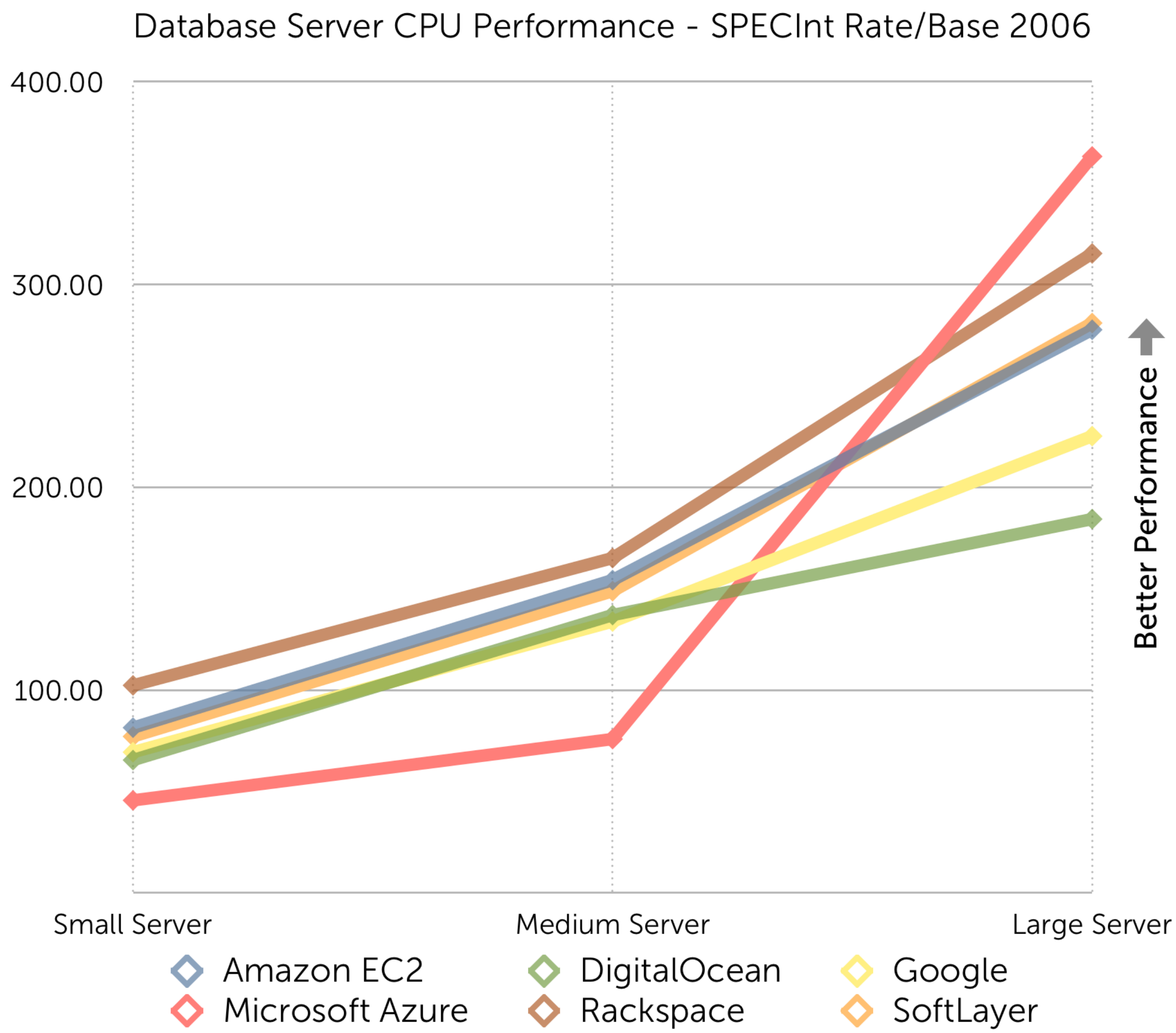
CPU Performance Variability

As with web server instances, CPU performance variability was observed both on the same instance with multiple benchmark runs, and across all runs from multiple instances. Because database server instances were larger, variability was generally lower. SoftLayer variability was due primarily to use of multiple CPU models for the same instance type. In order to compensate for such variability, the SPEC CPU 2006 metrics presented below are based on the mean minus standard deviation from all runs for each instance type.

CPU Performance Comments

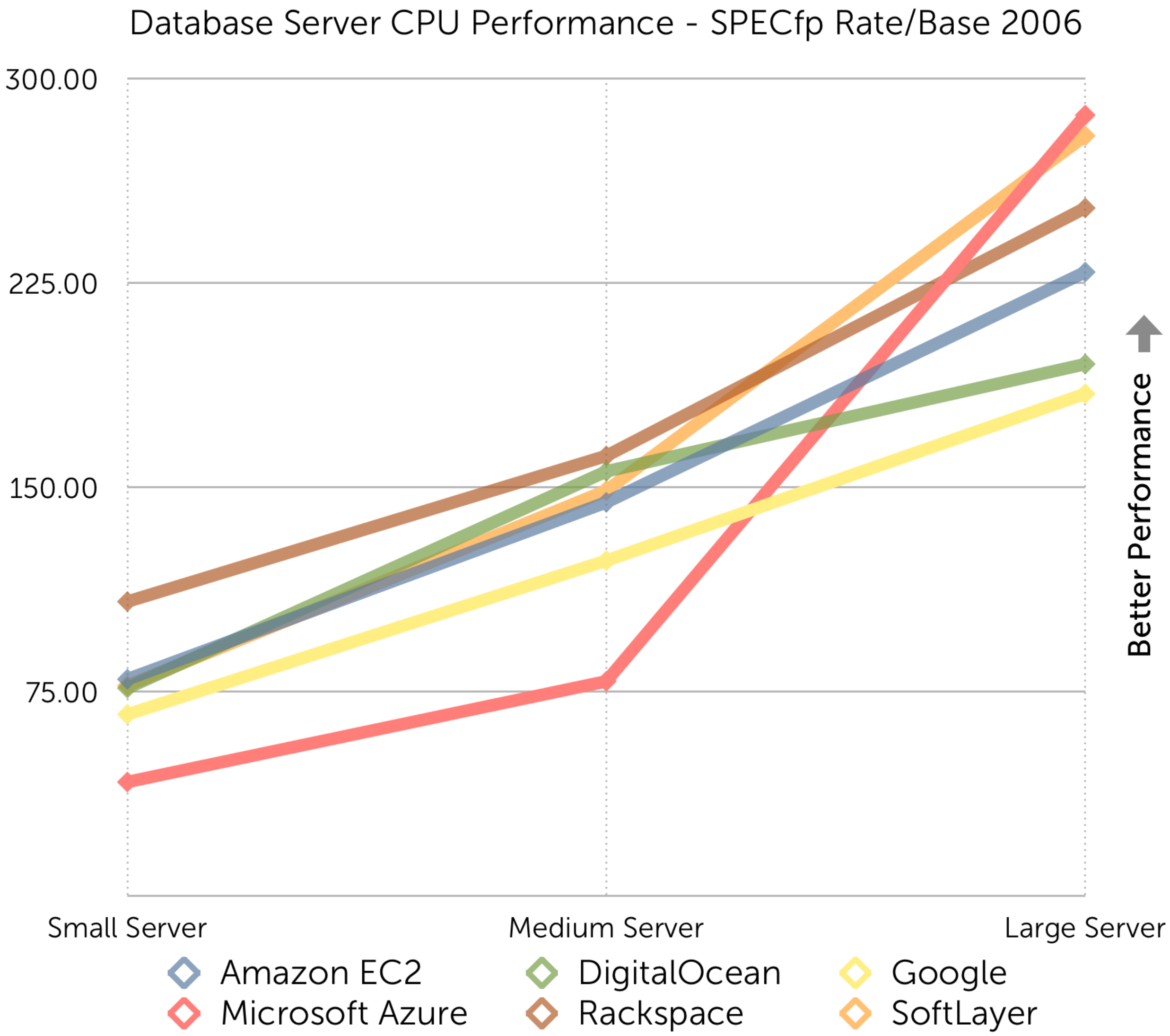
Overall database CPU performance results were similar to web server. Our SoftLayer compute instances again showed high CPU performance variability due to their use of multiple processor types. Rackspace Performance 2 database instances demonstrated much lower performance variability compared to the Performance 1 web server instances. Except for the large database server, Microsoft Azure results again lagged behind comparable compute instances because of its older AMD 4171 chipset.

SPECint Rate/Base 2006 (estimate)



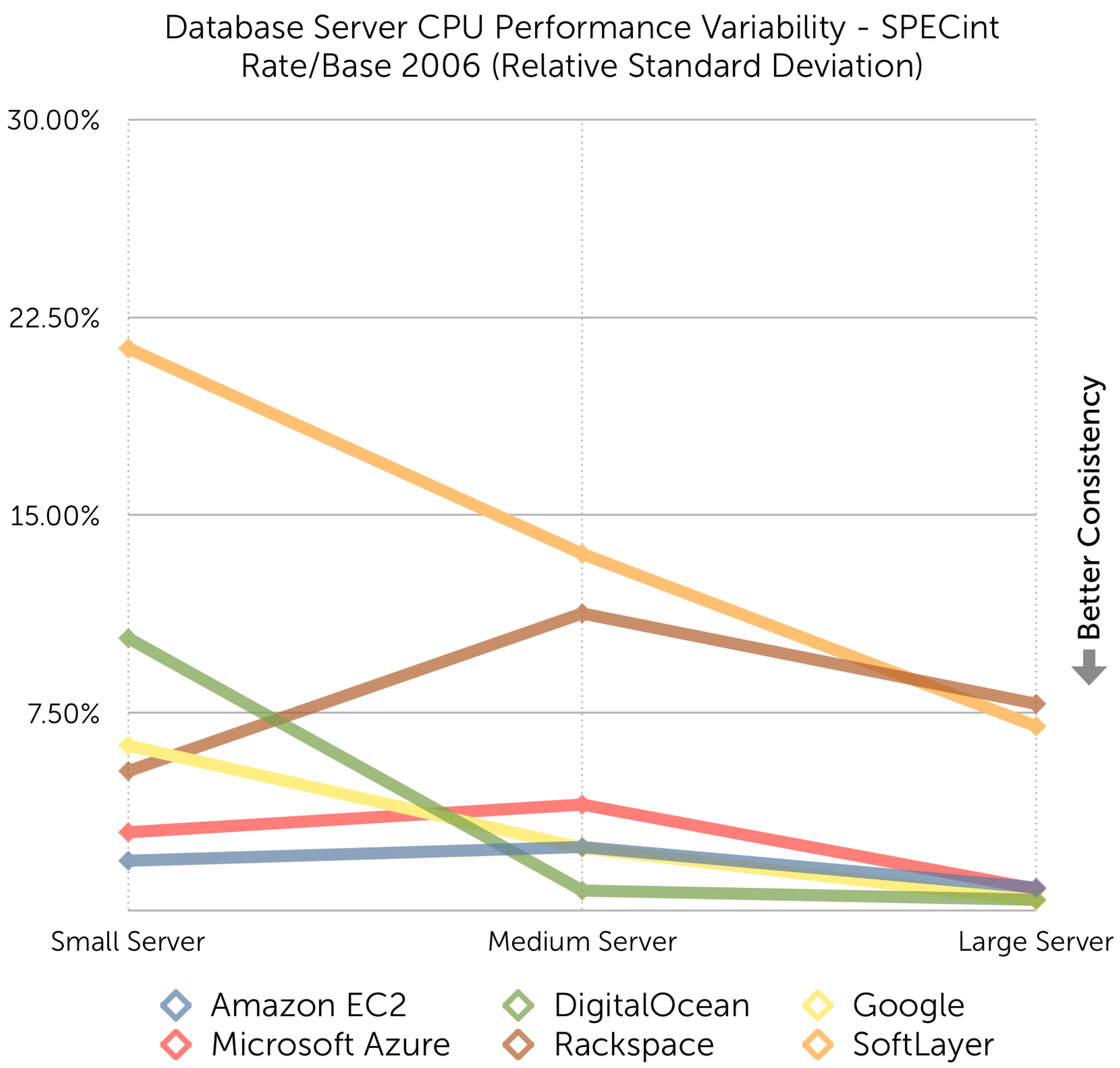
Server Type	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Server	81.43	65.46	69.37	45.60	106.26	77.11
Medium Server	154.19	136.95	133.75	75.82	182.89	148.73
Large Server	277.56	184.24	225.14	363.00	315.19	280.89

SPECfp Rate/Base 2006 (estimate)



Server Type	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Server	79.88	81.24	70.67	42.32	113.73	86.16
Medium Server	147.39	141.30	126.06	80.48	177.10	153.99
Large Server	230.92	195.52	184.45	286.63	272.65	259.59

CPU Performance Variability



Server Type	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Server	1.89%	10.33%	6.26%	2.97%	1.61%	21.32%
Medium Server	2.41%	0.76%	2.37%	4.02%	1.67%	13.53%
Large Server	0.87%	0.41%	0.38%	0.82%	7.84%	6.99%

Memory Performance

We used the STREAM benchmark to measure memory performance. STREAM is a simple synthetic benchmark program that measures sustainable memory bandwidth. The STREAM benchmark consists of four tests:

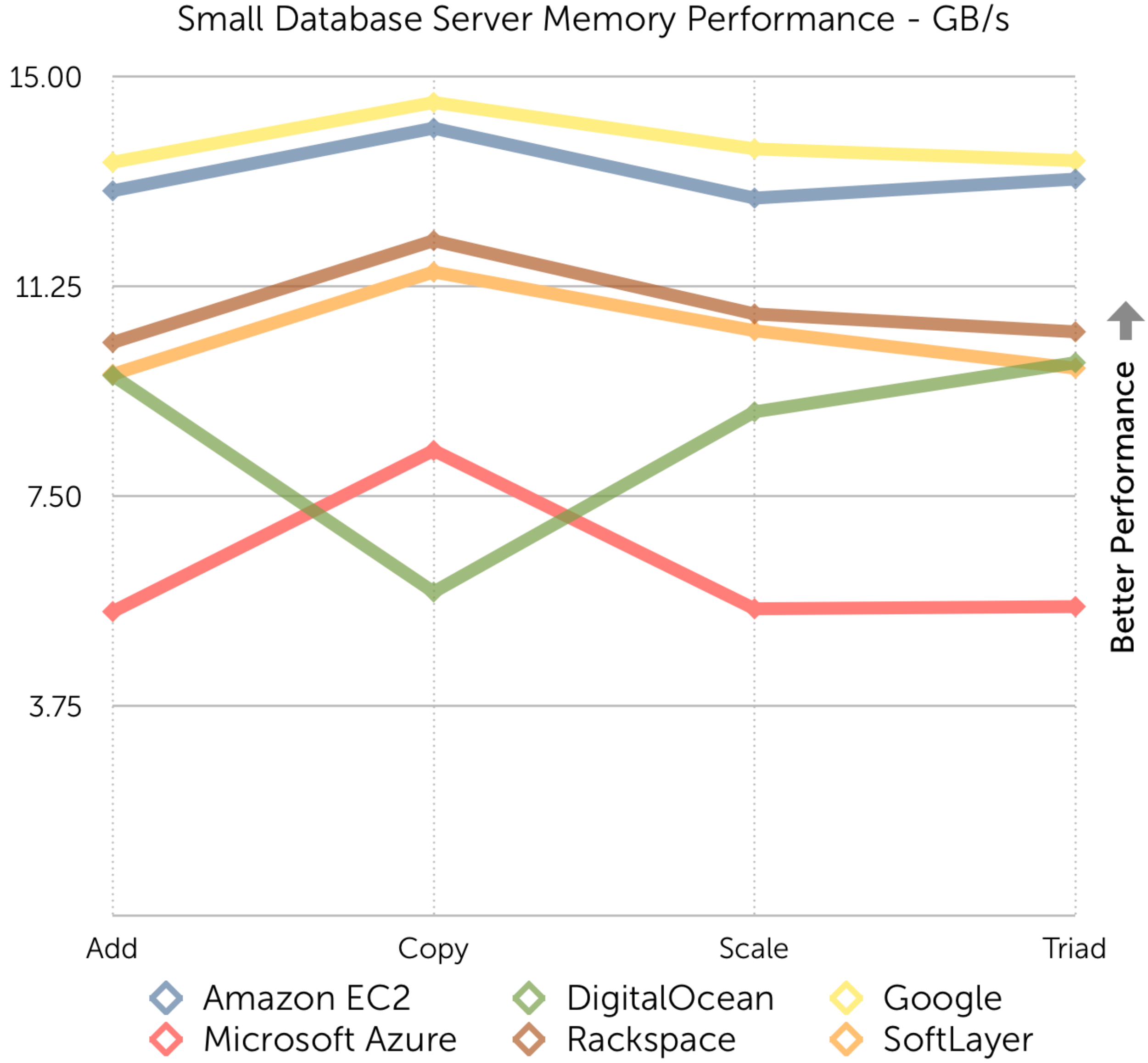
- *Copy* - measures transfer rates in the absence of arithmetic
- *Scale* - adds a simple arithmetic operation
- *Add* - adds a third operand to allow multiple load/store ports on vector machines to be tested
- *Triad* - allows chained/overlapped/fused multiply/add operations

The results presented represent sustainable bandwidth for each of these tests in gigabytes per second (GB/s), where a higher value represents better performance.

Memory Performance Comments

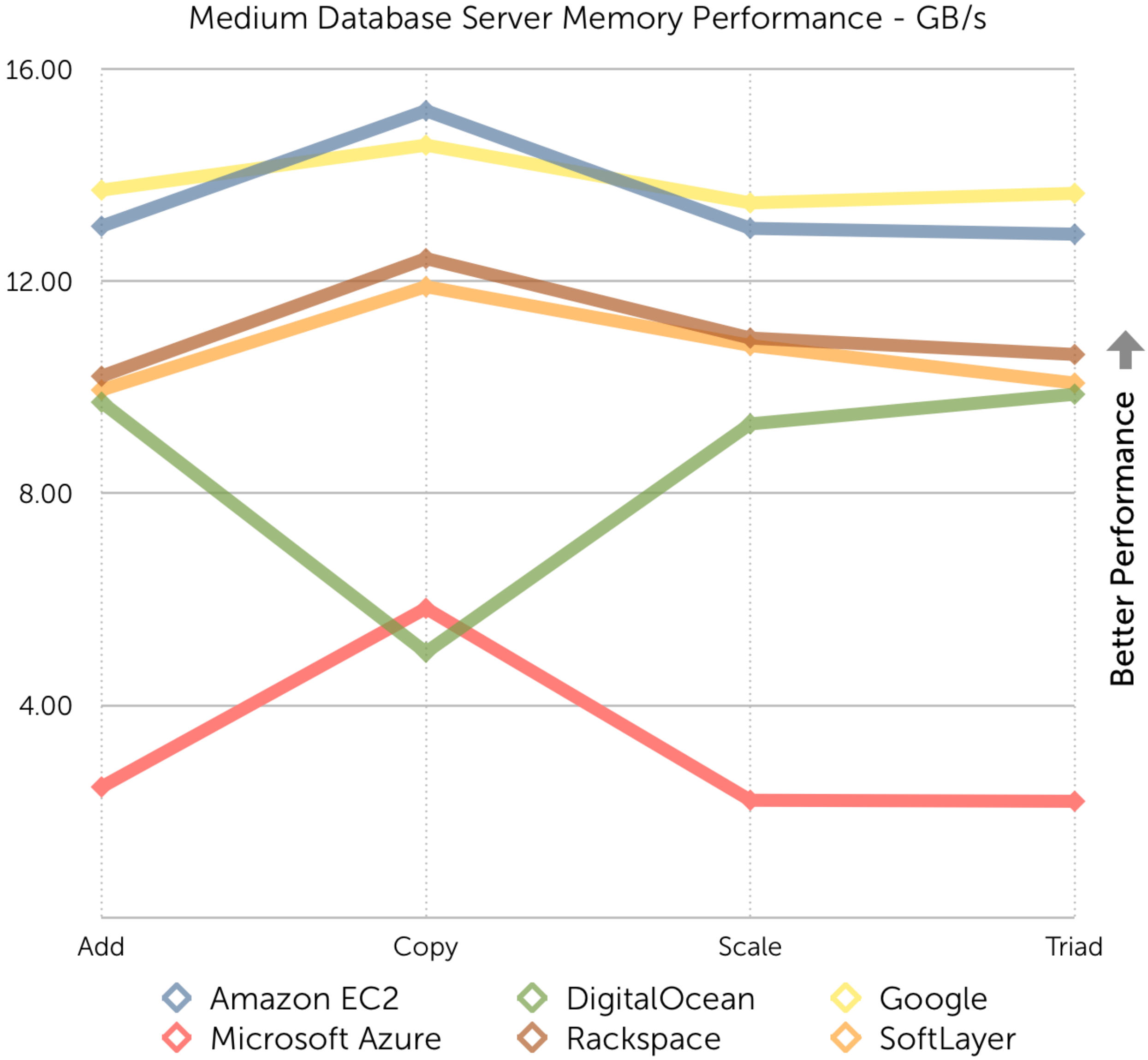
The limiting factor for memory performance is the hardware type and CPU chipset. Newer hardware generally supports faster memory and buses. Therefore, memory performance is generally similar across compute instances running on the same type of hardware, regardless of size. Microsoft Azure performed poorly in these tests for small and medium instance due to its use of older AMD 4171 chipsets (the large Azure instance uses Sandy Bridge). DigitalOcean results showed a consistent performance drop for memory copy tests. Amazon EC2 and Google consistently performed best for these tests.

Memory Performance - Small Database Server



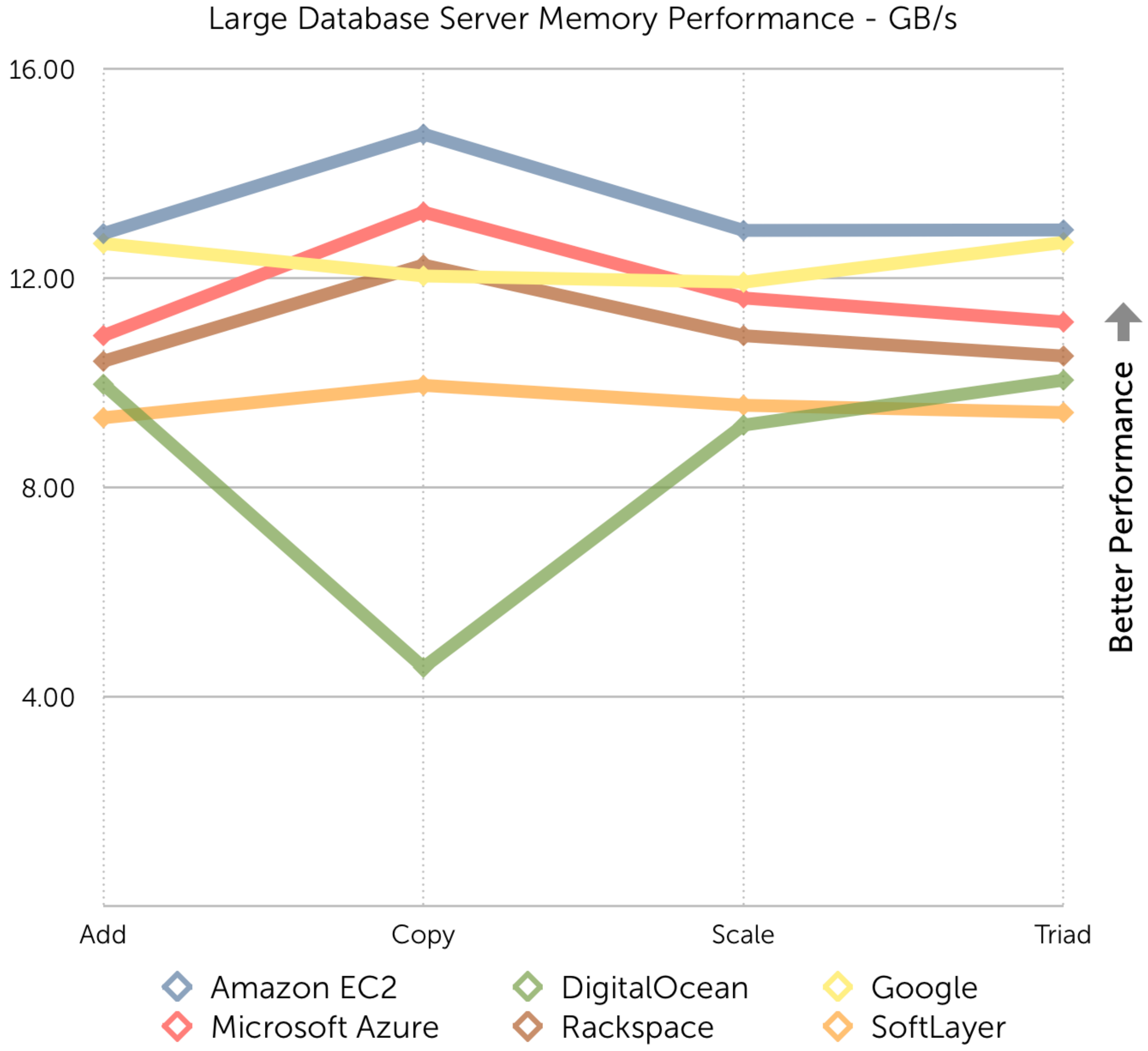
Server Type	Amazon EC2 (GB/s)	DigitalOcean (GB/s)	Google (GB/s)	Microsoft Azure (GB/s)	Rackspace (GB/s)	SoftLayer (GB/s)
Add	12.95	9.65	13.46	5.43	10.24	9.67
Copy	14.07	5.78	14.53	8.31	12.06	11.50
Scale	12.82	9.00	13.70	5.48	10.75	10.45
Triad	13.16	9.88	13.49	5.52	10.43	9.78

Memory Performance - Medium Database Server



Server Type	Amazon EC2 (GB/s)	DigitalOcean (GB/s)	Google (GB/s)	Microsoft Azure (GB/s)	Rackspace (GB/s)	SoftLayer (GB/s)
Add	13.04	9.72	13.72	2.46	10.21	9.95
Copy	15.22	5.00	14.57	5.83	12.43	11.90
Scale	13.00	9.31	13.48	2.21	10.93	10.79
Triad	12.89	9.87	13.66	2.19	10.62	10.08

Memory Performance - Large Database Server



Server Type	Amazon EC2 (GB/s)	DigitalOcean (GB/s)	Google (GB/s)	Microsoft Azure (GB/s)	Rackspace (GB/s)	SoftLayer (GB/s)
Add	12.85	9.97	12.66	10.90	10.41	9.33
Copy	14.75	4.56	12.04	13.26	12.25	9.95
Scale	12.91	9.19	11.92	11.62	10.90	9.57
Triad	12.92	10.05	12.68	11.16	10.51	9.43

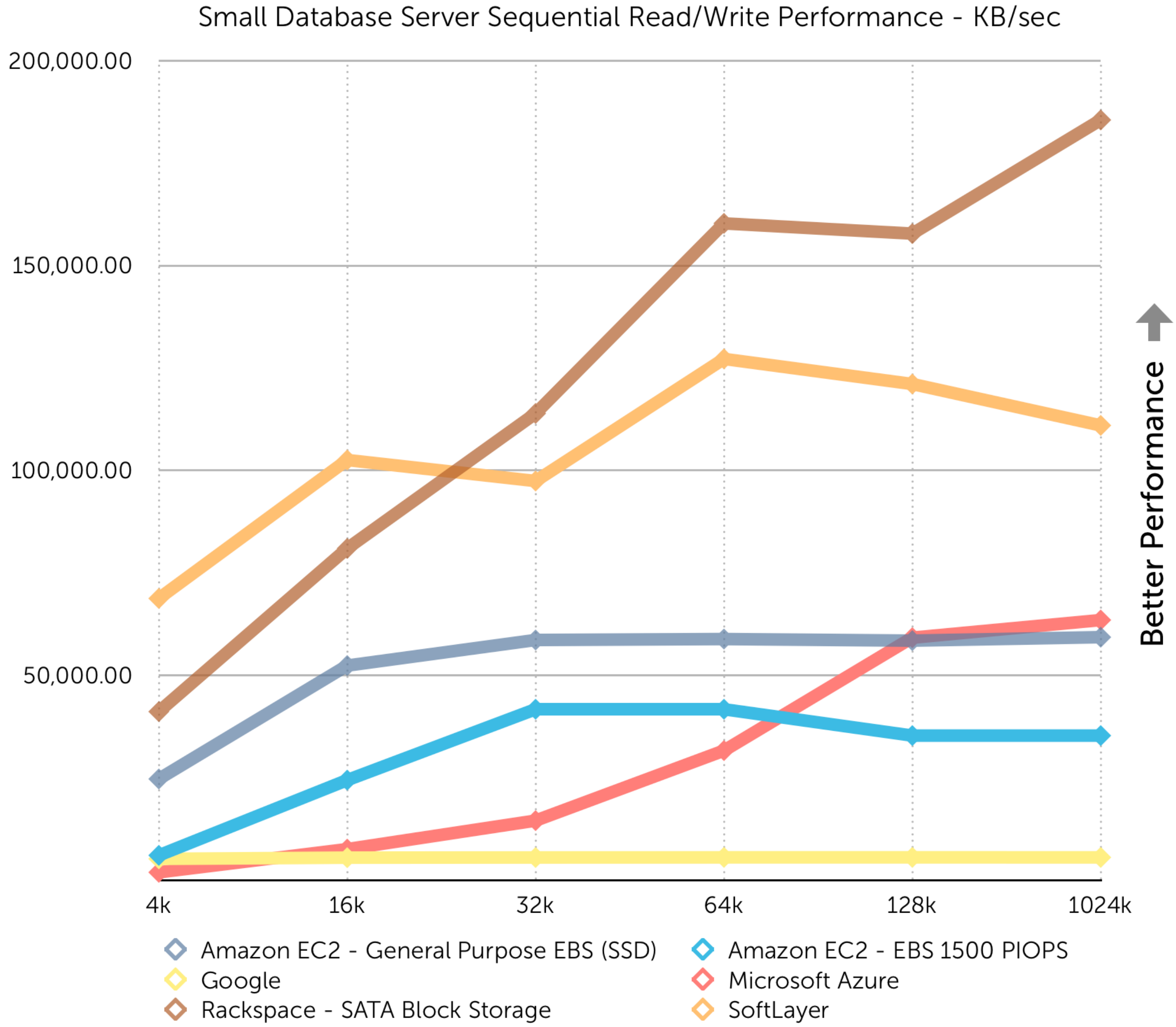
Disk Performance

We measured disk performance for the database servers using *fio*. The focus of the database server results provided below is two sets of read/write workloads (80-percent read, 20-percent write), sequential and random. These workloads might approximate a database server performing both read and write queries. Our testing was against a single disk volume. You may be able to achieve higher throughput by combining multiple volumes with software Raid. Unlike the web server analysis, which focused on local storage, for database servers the focus was on external storage. External storage typically involves connecting to disk volumes over a network as opposed to using local storage that is directly attached to host hardware. Because external storage disk operations traverse a network, performance variability is often much higher. This variability is reflected in the performance consistency graphs and tables below.

Disk Performance Comments

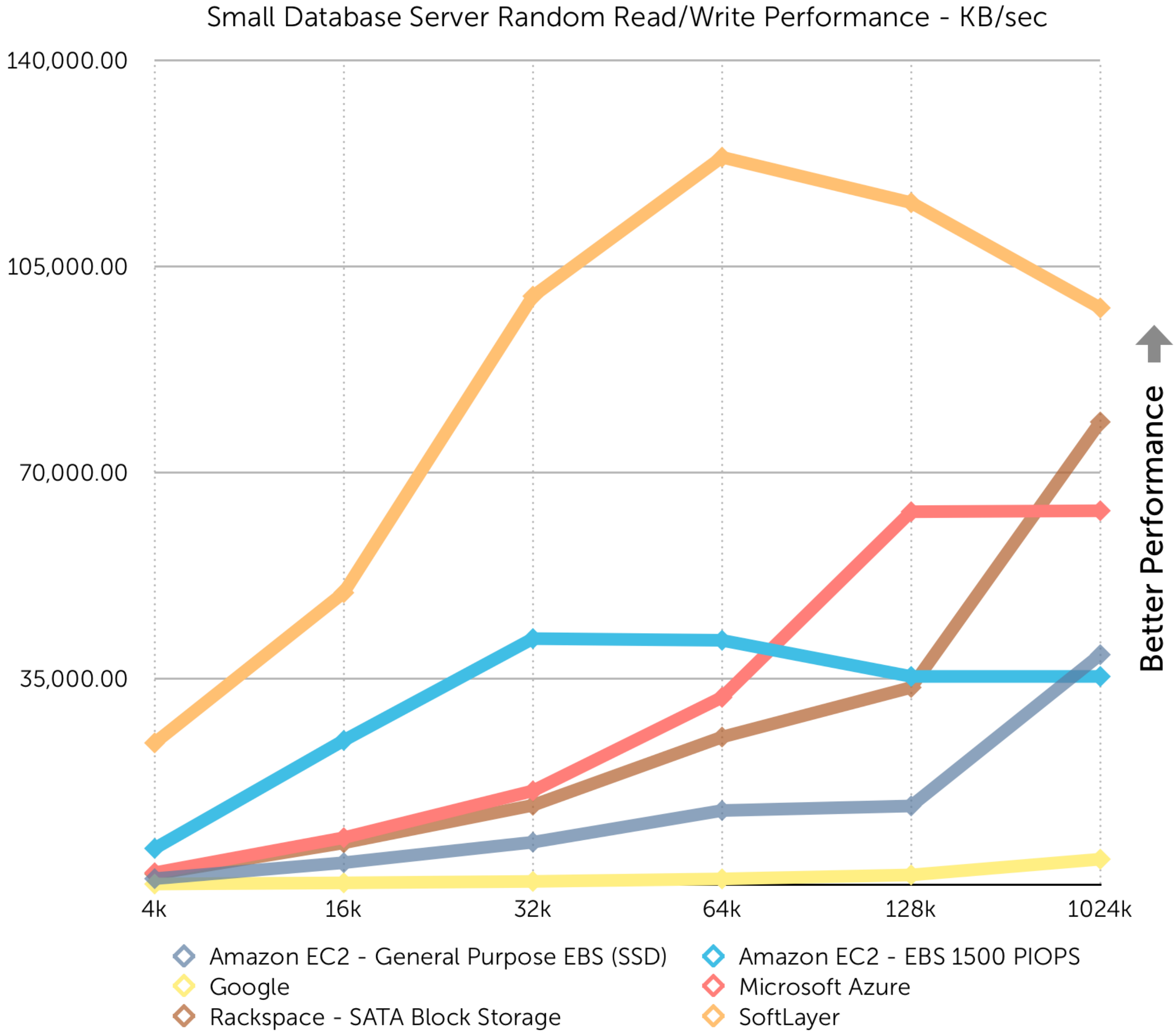
Amazon EBS provisioned IOPS (PIOPS) volumes performance was highly consistent in all tests, more so than any other service, including General Purpose EBS (SSD). This is because the EBS PIOPS approach is very different from most external storage platforms because you get exactly the amount of IO you provision. Other EBS test results are based on current General Purpose (SSD) based volumes (referred to as EBS in the proceeding graphs). We excluded DigitalOcean from the test results because they do not offer external storage. Google IO performance scaled linearly by instance size, but performance was generally slower. Rackspace SSD and SATA Cloud Block Storage performed faster than any other service, but also had high performance variability (likely due to use of a burstable storage network). SoftLayer also performed well, but demonstrated high performance variability.

Sequential Read/Write Disk Performance - Small Database Server



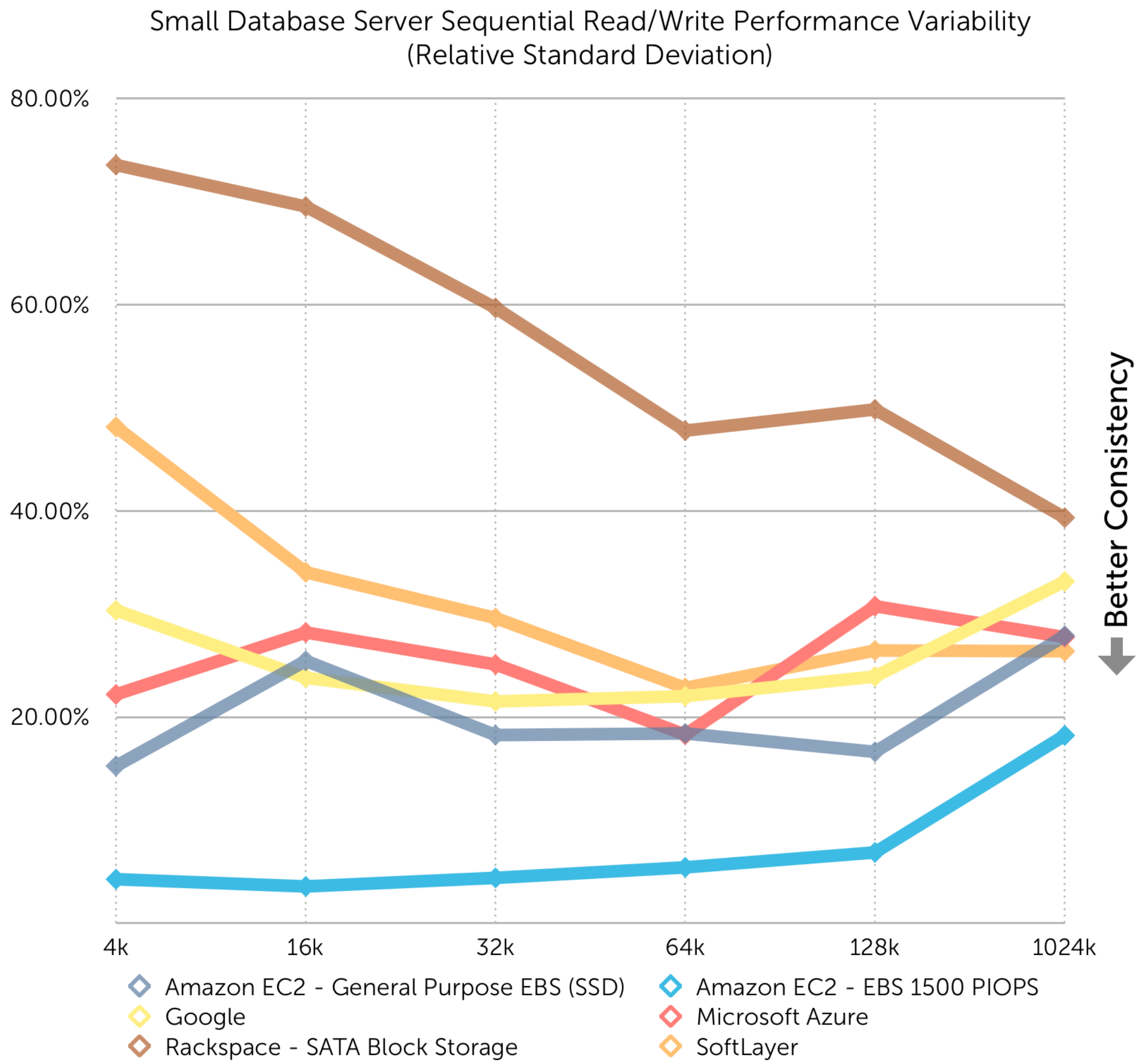
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 1500 PIOPS (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace - SATA Block Storage (KB/s)	SoftLayer (KB/s)
4k	24,756.00	6,117.67	5,264.67	1,977.33	41,184.67	68,794.00
16k	52,415.33	24,474.00	5,557.00	7,644.33	80,988.67	102,566.33
32k	58,659.67	41,766.67	5,596.67	14,594.33	113,923.00	97,434.00
64k	58,888.33	41,767.33	5,629.33	31,676.00	160,290.67	127,230.33
128k	58,501.00	35,333.00	5,630.00	59,221.33	157,793.33	121,104.33
1024k	59,342.33	35,334.00	5,620.00	63,526.33	185,523.33	110,998.67

Random Read/Write Disk Performance - Small Database Server



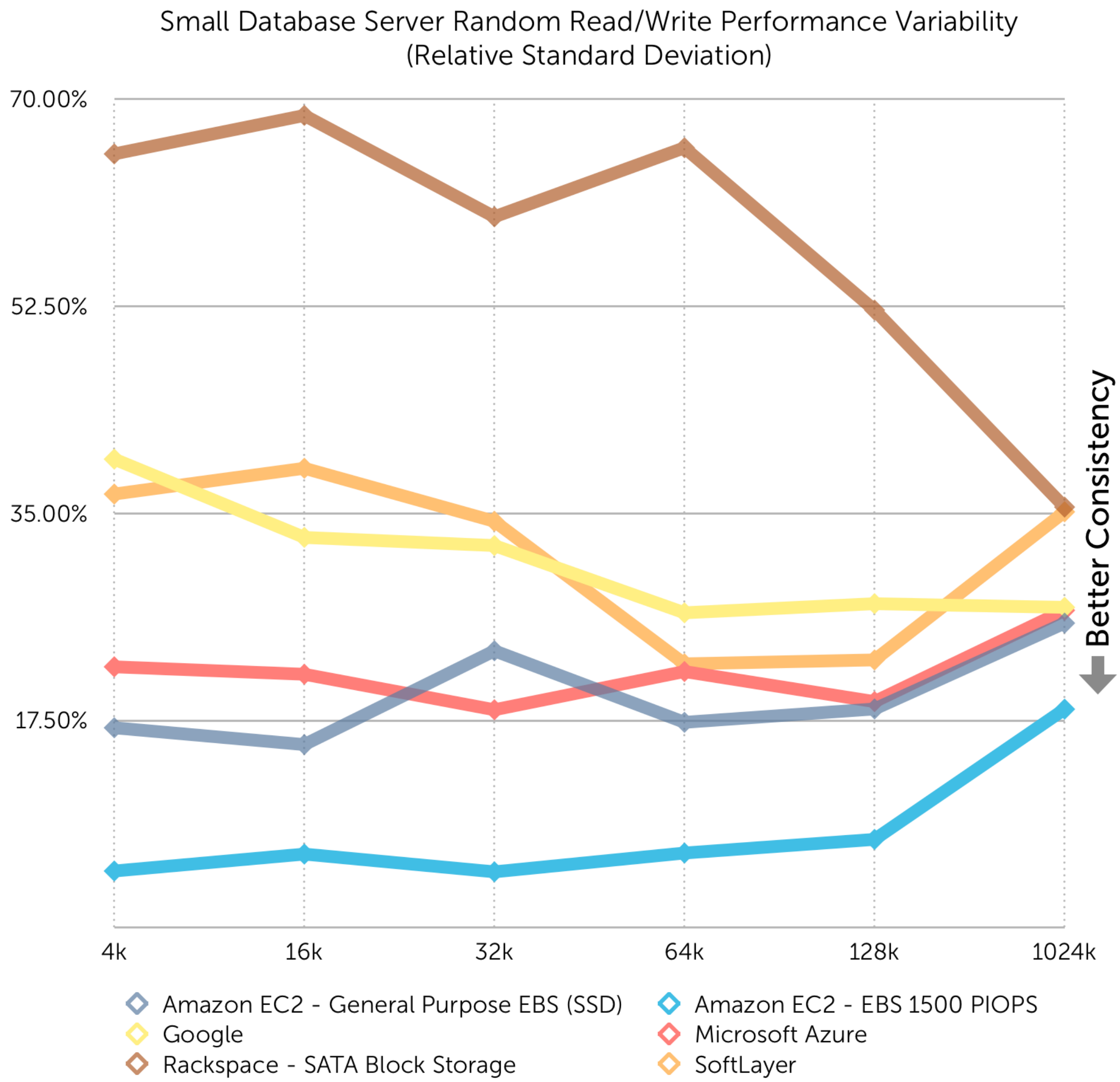
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 1500 PIOPS (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace - SATA Block Storage (KB/s)	SoftLayer (KB/s)
4k	978.00	6,120.67	70.00	1,983.67	1,788.00	24,045.67
16k	3,693.00	24,469.33	267.00	7,936.00	7,061.67	49,572.33
32k	7,170.67	41,766.33	511.00	15,896.00	13,448.67	99,952.33
64k	12,556.67	41,478.67	938.33	31,784.67	25,041.00	123,564.33
128k	13,332.00	35,334.00	1,610.67	63,325.67	33,465.67	115,835.33
1024k	39,039.67	35,333.67	4,290.00	63,493.33	78,577.00	97,954.33

Sequential Disk Read/Write Performance Consistency - Small Database Server



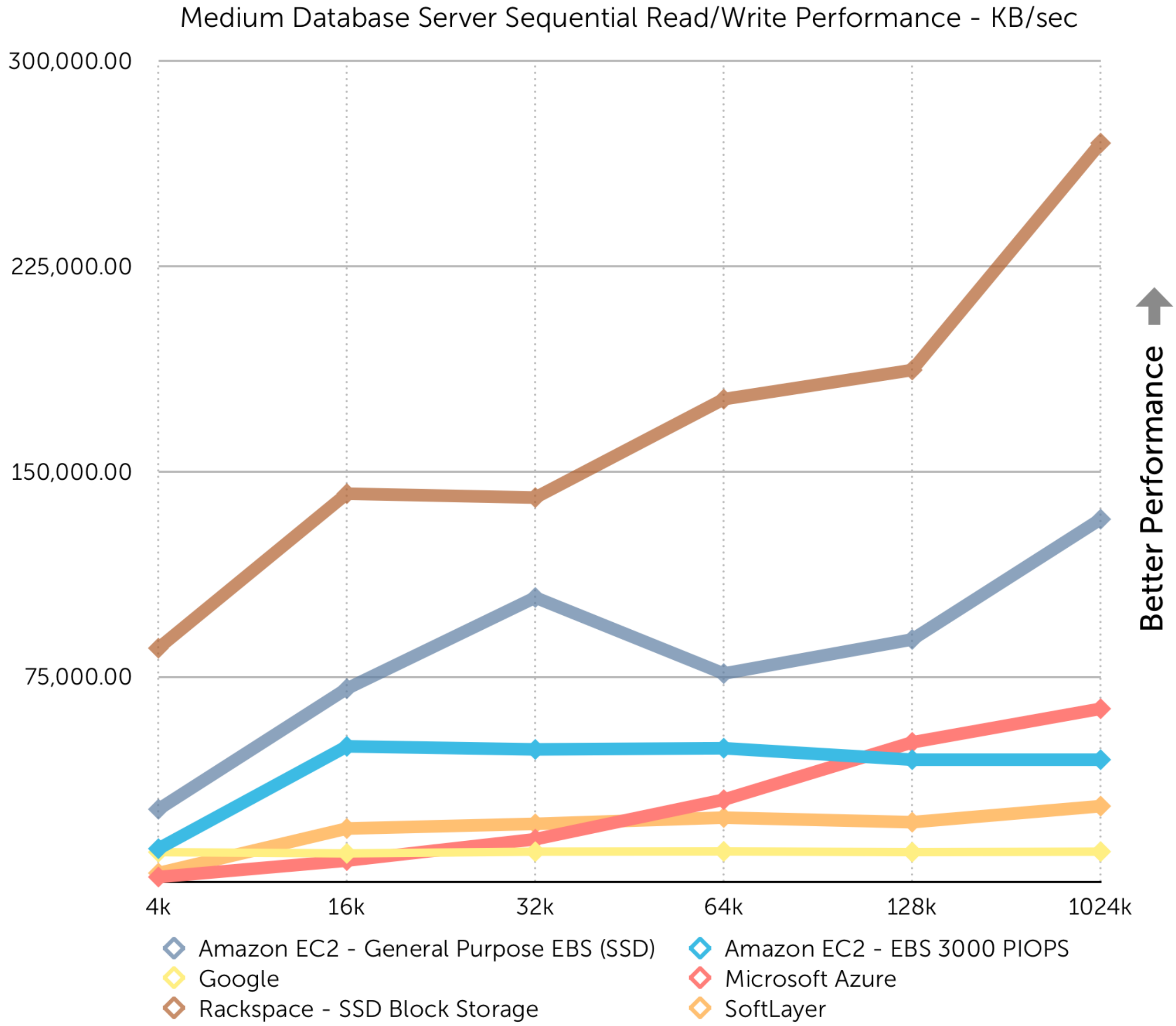
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 1500 PIOPS	Google	Microsoft Azure	Rackspace - SATA Block Storage	SoftLayer
4k	15.26%	4.28%	30.36%	22.20%	73.54%	48.14%
16k	25.46%	3.58%	23.81%	28.21%	69.51%	34.05%
32k	18.26%	4.41%	21.52%	25.11%	59.68%	29.61%
64k	18.42%	5.43%	22.03%	18.30%	47.78%	22.83%
128k	16.63%	6.88%	23.94%	30.76%	49.84%	26.46%
1024k	27.89%	18.22%	33.16%	27.75%	39.34%	26.39%

Random Disk Read/Write Performance Consistency - Small Database Server



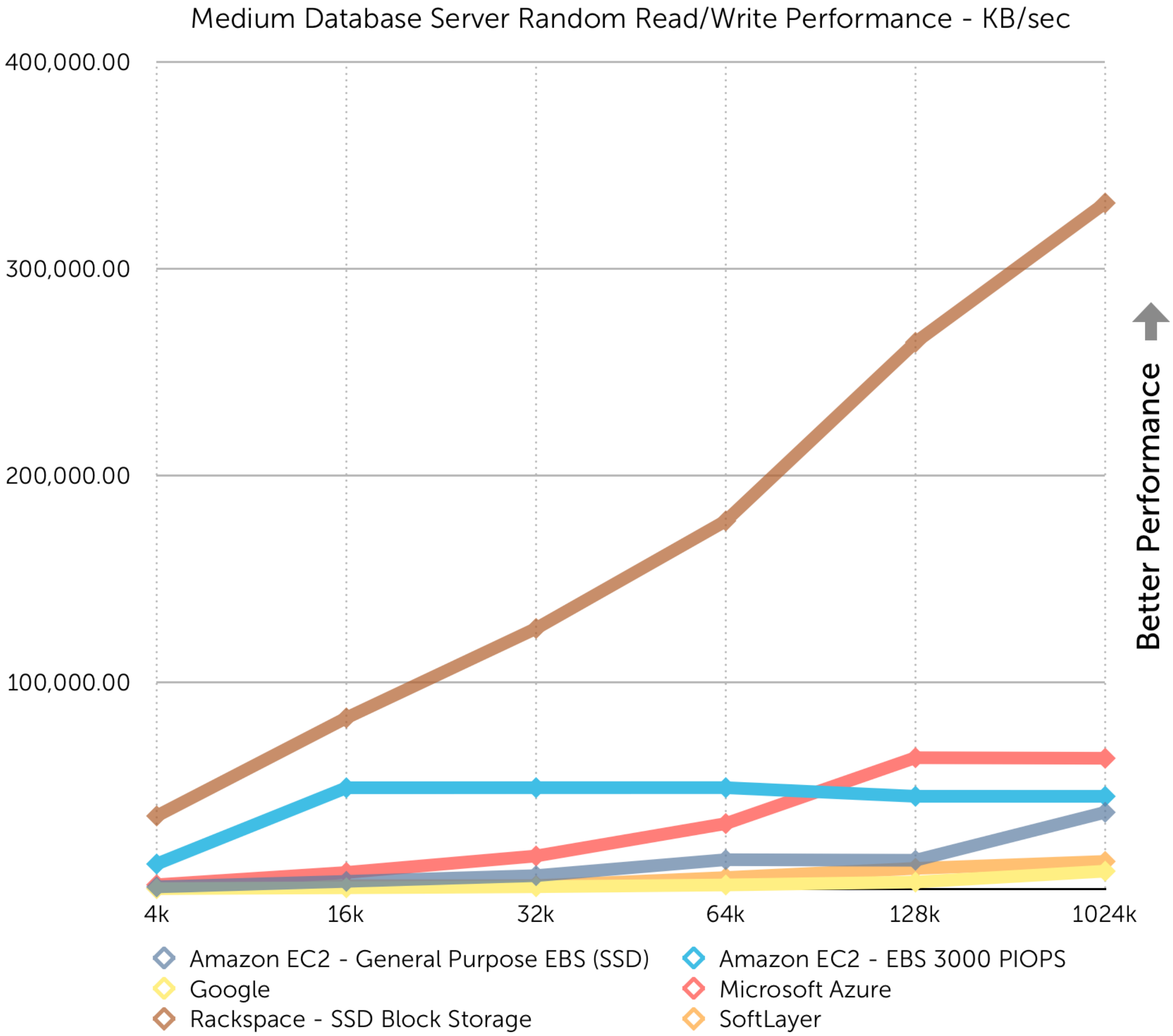
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 1500 PIOPS	Google	Microsoft Azure	Rackspace - SATA Block Storage	SoftLayer
4k	16.89%	4.79%	39.58%	22.04%	65.36%	36.63%
16k	15.47%	6.22%	32.97%	21.41%	68.61%	38.81%
32k	23.39%	4.72%	32.28%	18.41%	60.05%	34.33%
64k	17.35%	6.31%	26.60%	21.64%	65.88%	22.29%
128k	18.48%	7.46%	27.37%	19.10%	52.17%	22.62%
1024k	25.73%	18.46%	27.06%	26.81%	35.53%	35.13%

Sequential Read/Write Disk Performance - Medium Database Server



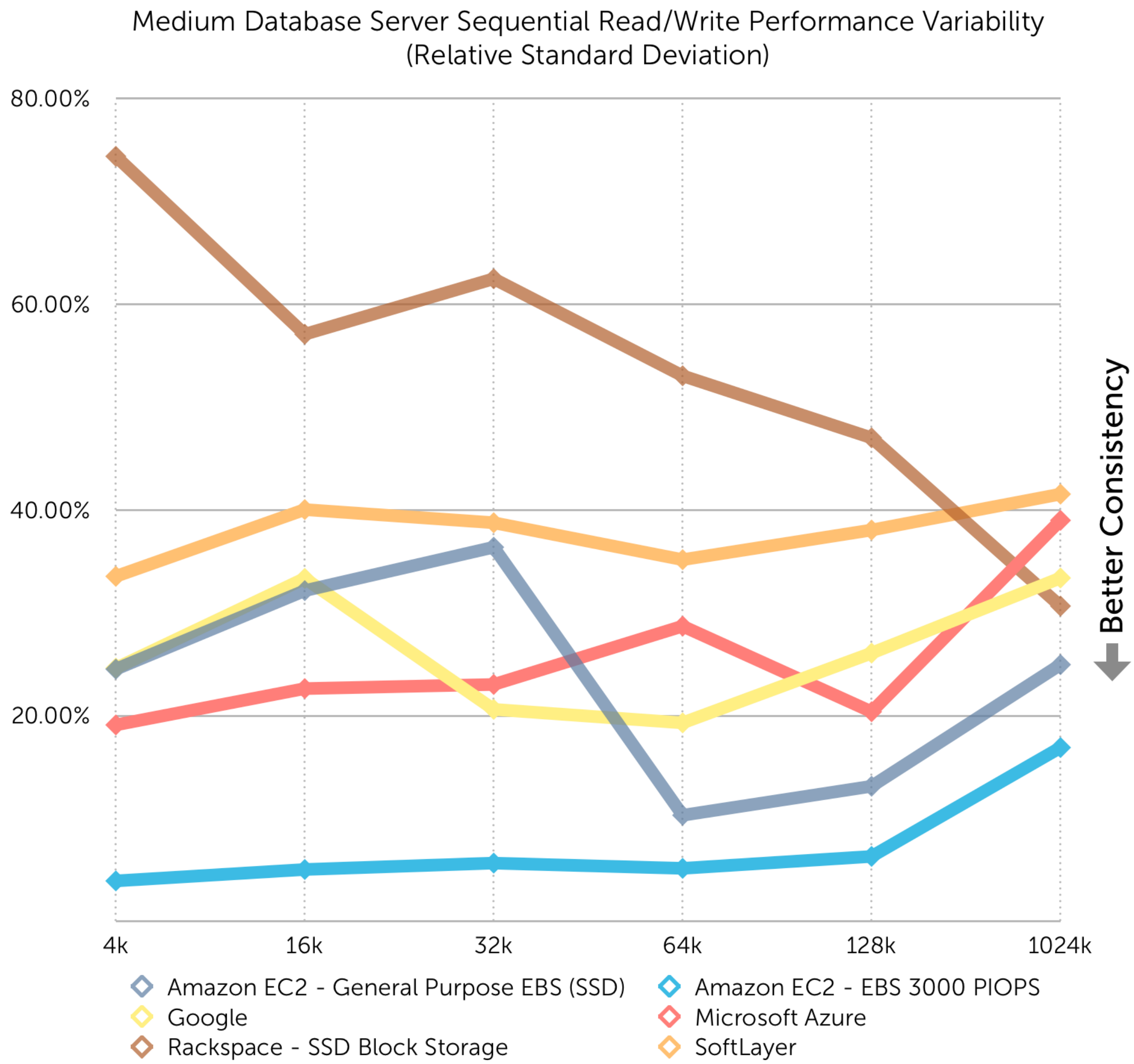
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 3000 PIOPS (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace - SSD Block Storage (KB/s)	SoftLayer (KB/s)
4k	26,650.00	12,260.67	10,990.33	1,905.67	85,503.67	3,341.33
16k	70,712.33	49,632.67	10,528.67	7,909.67	141,920.00	19,613.67
32k	104,004.33	48,515.00	11,168.00	15,714.67	140,496.67	21,315.33
64k	76,210.00	48,952.00	11,262.33	30,151.00	176,448.67	23,608.00
128k	88,676.00	44,757.00	10,998.00	51,085.67	187,089.33	21,930.00
1024k	132,608.00	44,756.67	11,182.67	63,373.33	269,966.00	27,792.67

Random Read/Write Disk Performance - Medium Database Server



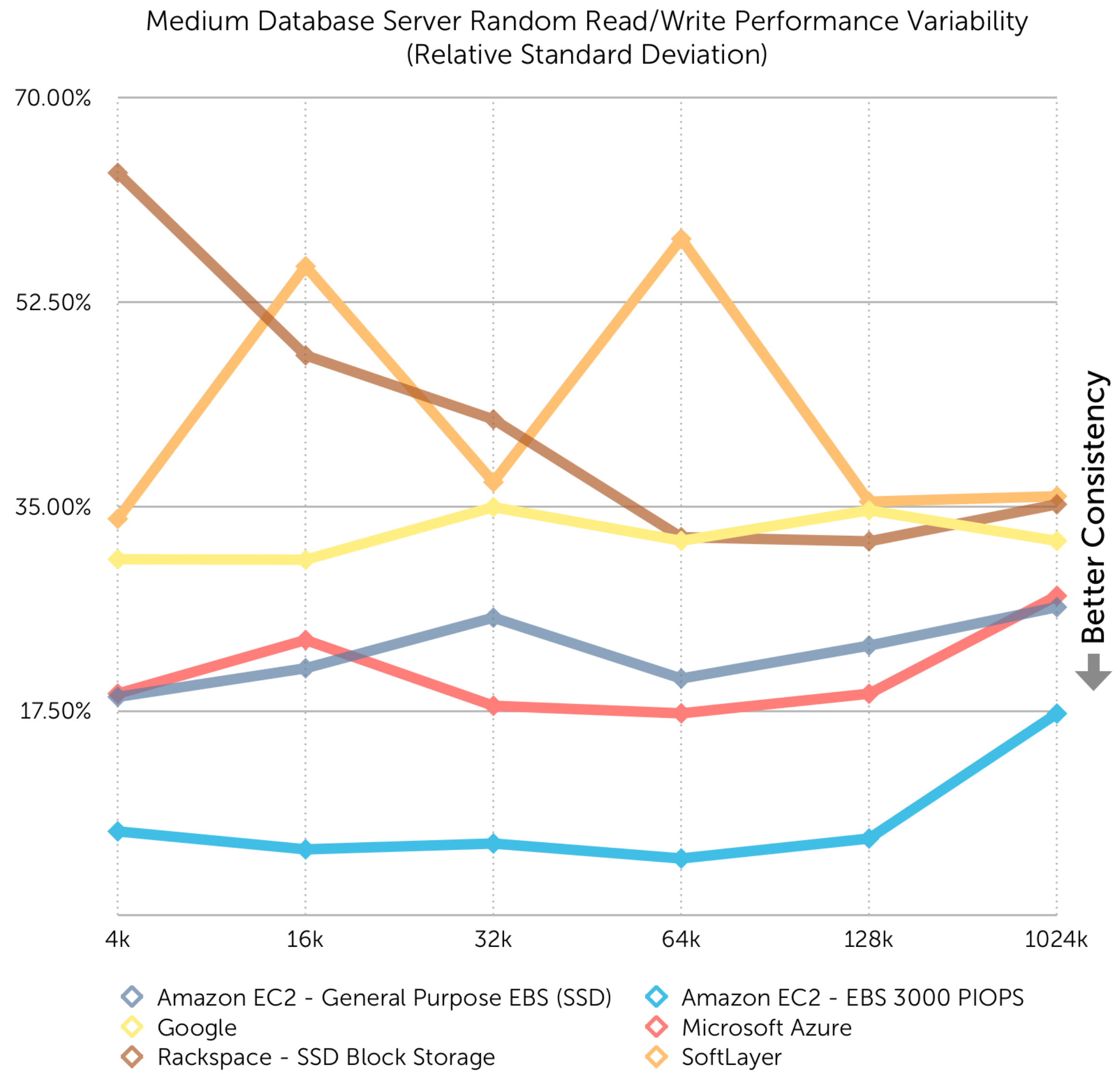
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 3000 PIOPS (KB/s)	Google (KB/s)	Microsoft Azure (KB/s)	Rackspace - SSD Block Storage (KB/s)	SoftLayer (KB/s)
4k	958.33	12,035.00	137.33	1,987.67	35,273.00	208.33
16k	3,638.00	48,918.33	535.00	7,952.33	82,732.00	1,515.00
32k	6,401.67	48,956.67	1,040.67	15,908.67	126,008.67	1,532.00
64k	14,125.67	48,967.67	1,906.67	31,592.67	178,040.00	5,470.67
128k	13,840.00	44,783.67	3,285.33	63,468.67	264,462.33	9,910.00
1024k	37,042.33	44,768.67	8,579.67	63,188.67	331,863.67	13,275.00

Sequential Disk Read/Write Performance Consistency - Medium Database Server



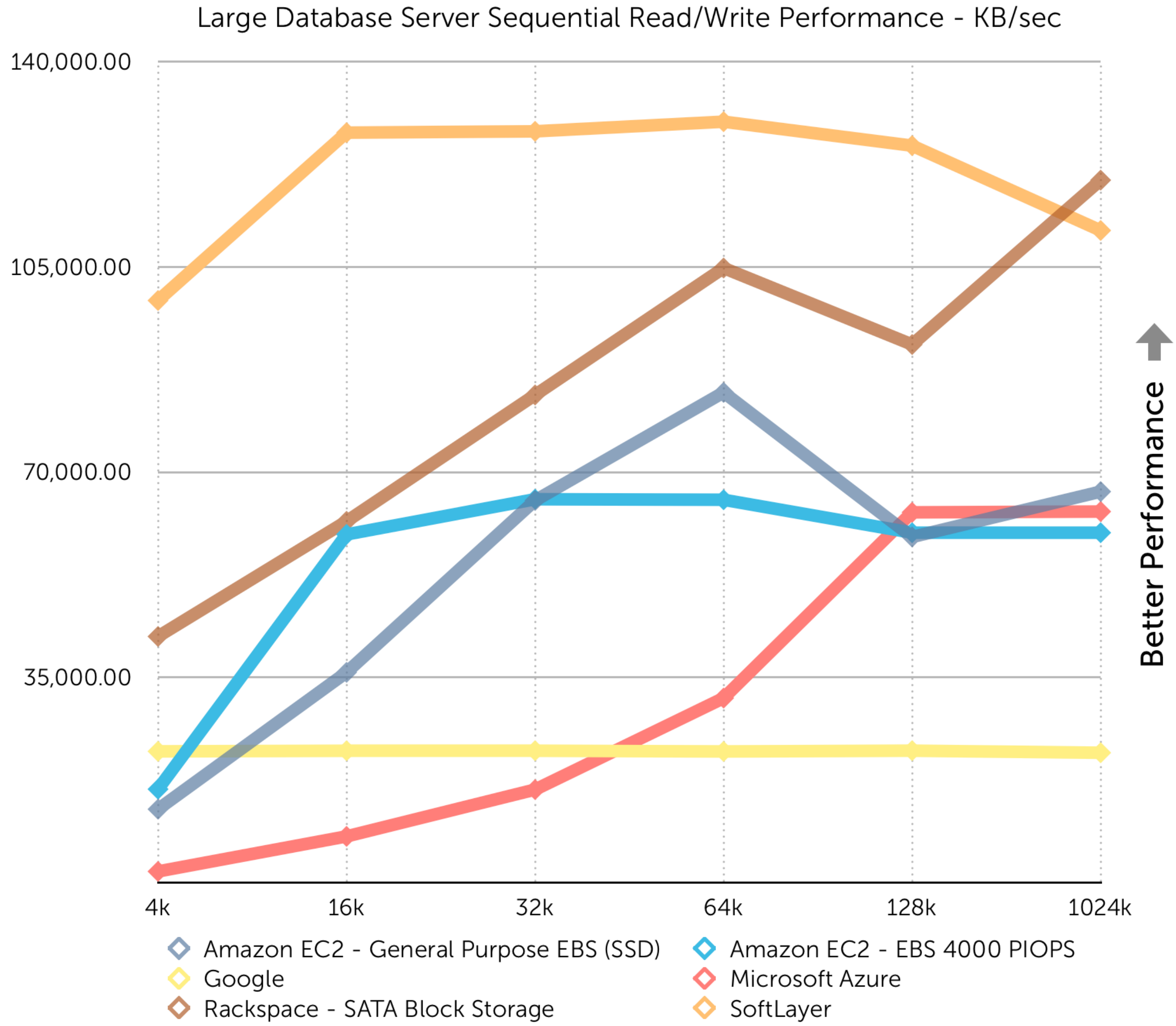
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 3000 PIOPS	Google	Microsoft Azure	Rackspace - SSD Block Storage	SoftLayer
4k	24.53%	3.95%	24.63%	19.11%	74.37%	33.55%
16k	32.16%	5.06%	33.35%	22.64%	57.08%	40.04%
32k	36.38%	5.68%	20.62%	23.04%	62.47%	38.76%
64k	10.33%	5.17%	19.31%	28.71%	53.04%	35.17%
128k	13.15%	6.33%	26.09%	20.39%	47.01%	38.05%
1024k	24.97%	16.91%	33.39%	38.99%	30.66%	41.54%

Random Disk Read/Write Performance Consistency - Medium Database Server



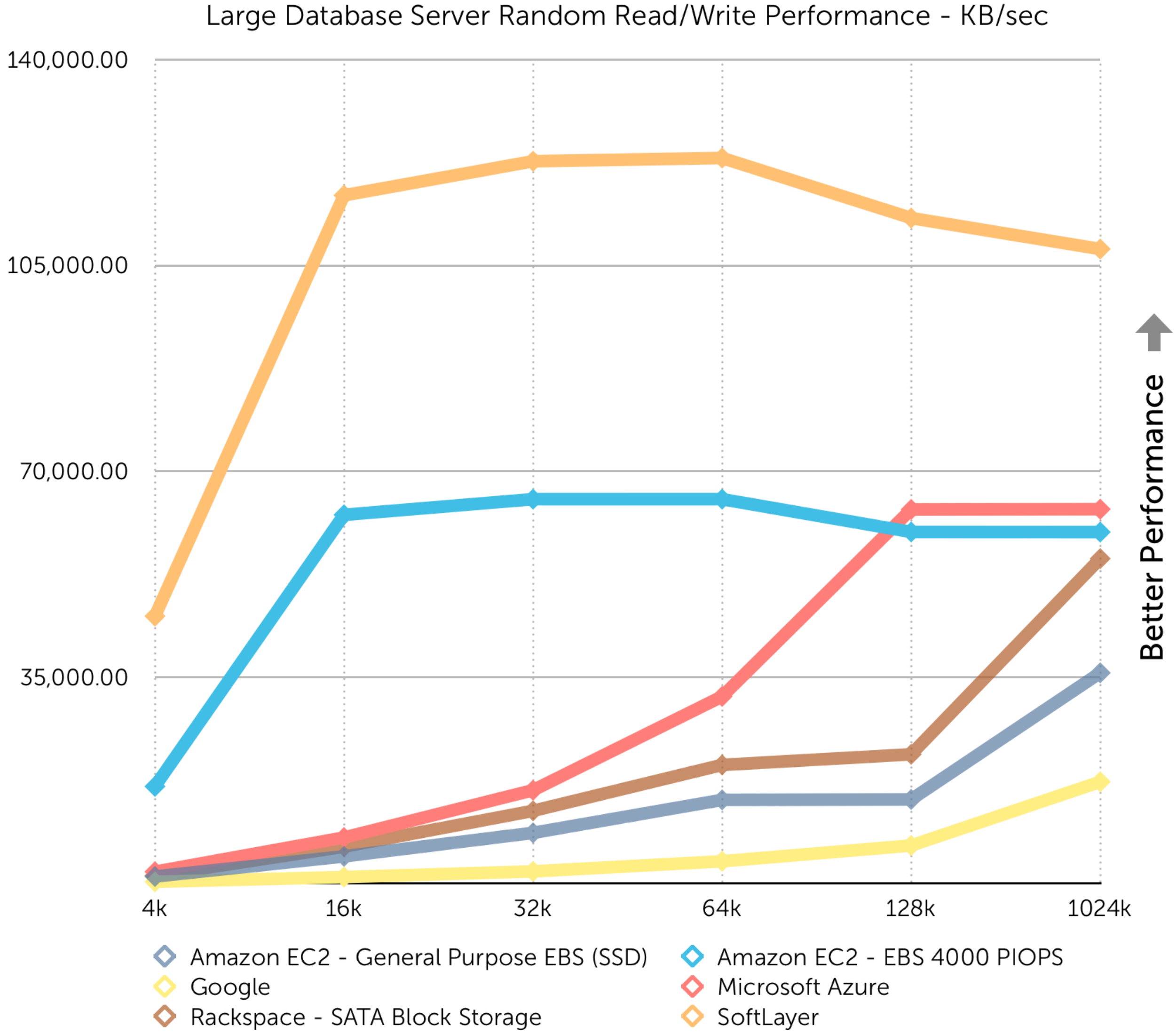
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 3000 PIOPS	Google	Microsoft Azure	Rackspace - SSD Block Storage	SoftLayer
4k	18.71%	7.20%	30.49%	18.99%	63.56%	33.95%
16k	21.16%	5.66%	30.45%	23.58%	47.94%	55.56%
32k	25.49%	6.16%	34.95%	17.95%	42.46%	37.09%
64k	20.28%	4.88%	32.07%	17.30%	32.38%	57.91%
128k	23.10%	6.59%	34.68%	18.98%	32.02%	35.42%
1024k	26.39%	17.28%	32.06%	27.36%	35.19%	35.88%

Sequential Read/Write Disk Performance - Large Database Server



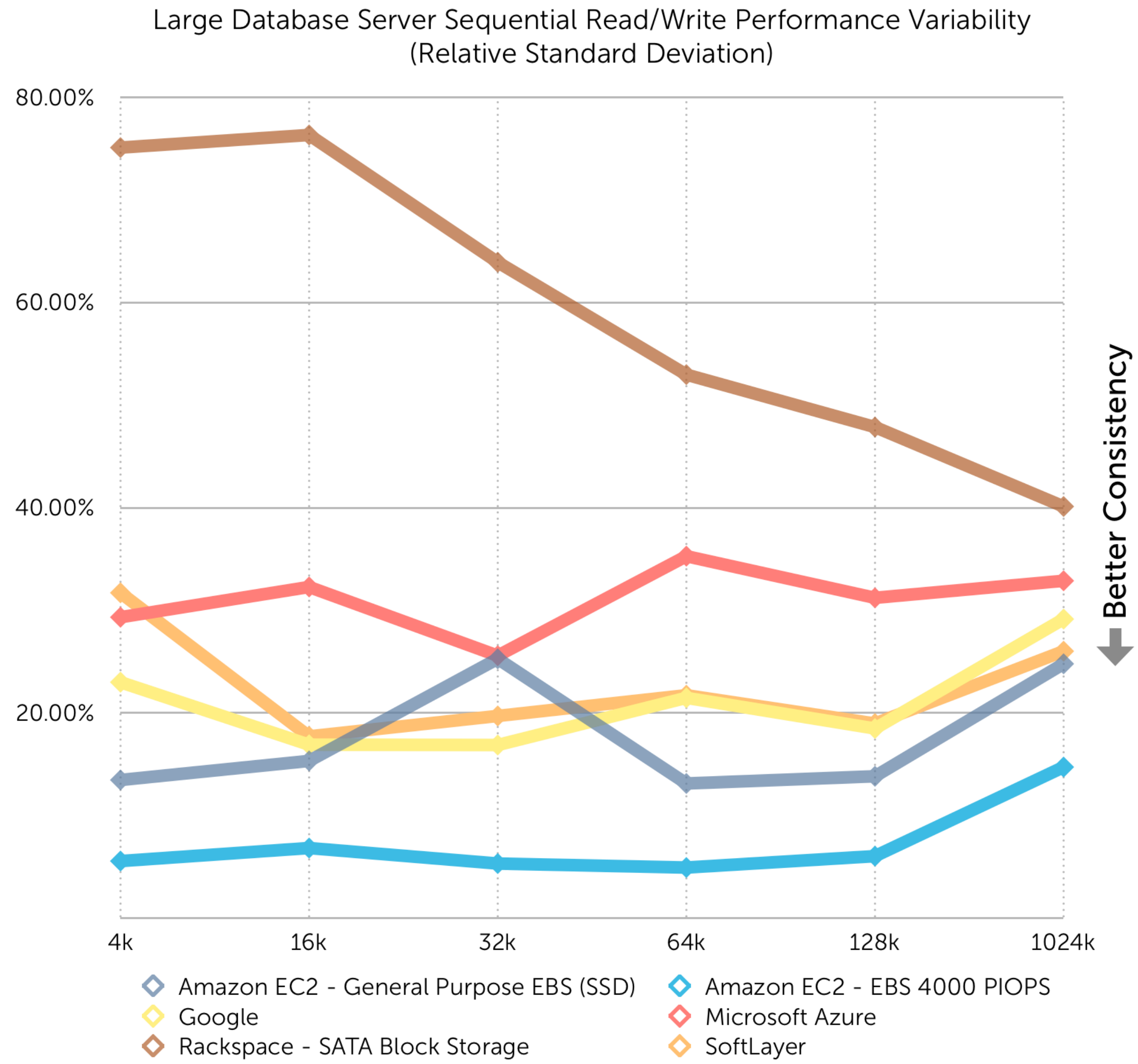
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 4000 PIOPS (KB/s)	Google (KB/s)	Microsoft Azure	Rackspace - SATA Block Storage (KB/s)	SoftLayer (KB/s)
4k	12,549.67	15,964.67	22,414.33	NA	41,992.67	99,259.67
16k	35,874.00	59,384.33	22,516.00	NA	61,623.00	127,871.67
32k	65,170.67	65,388.67	22,519.67	NA	83,157.67	128,092.67
64k	83,594.00	65,297.00	22,397.00	NA	104,796.33	129,739.33
128k	58,949.00	59,680.33	22,530.00	NA	91,751.00	125,649.67
1024k	66,685.33	59,677.67	22,177.00	NA	119,773.33	111,194.33

Random Read/Write Disk Performance - Large Database Server



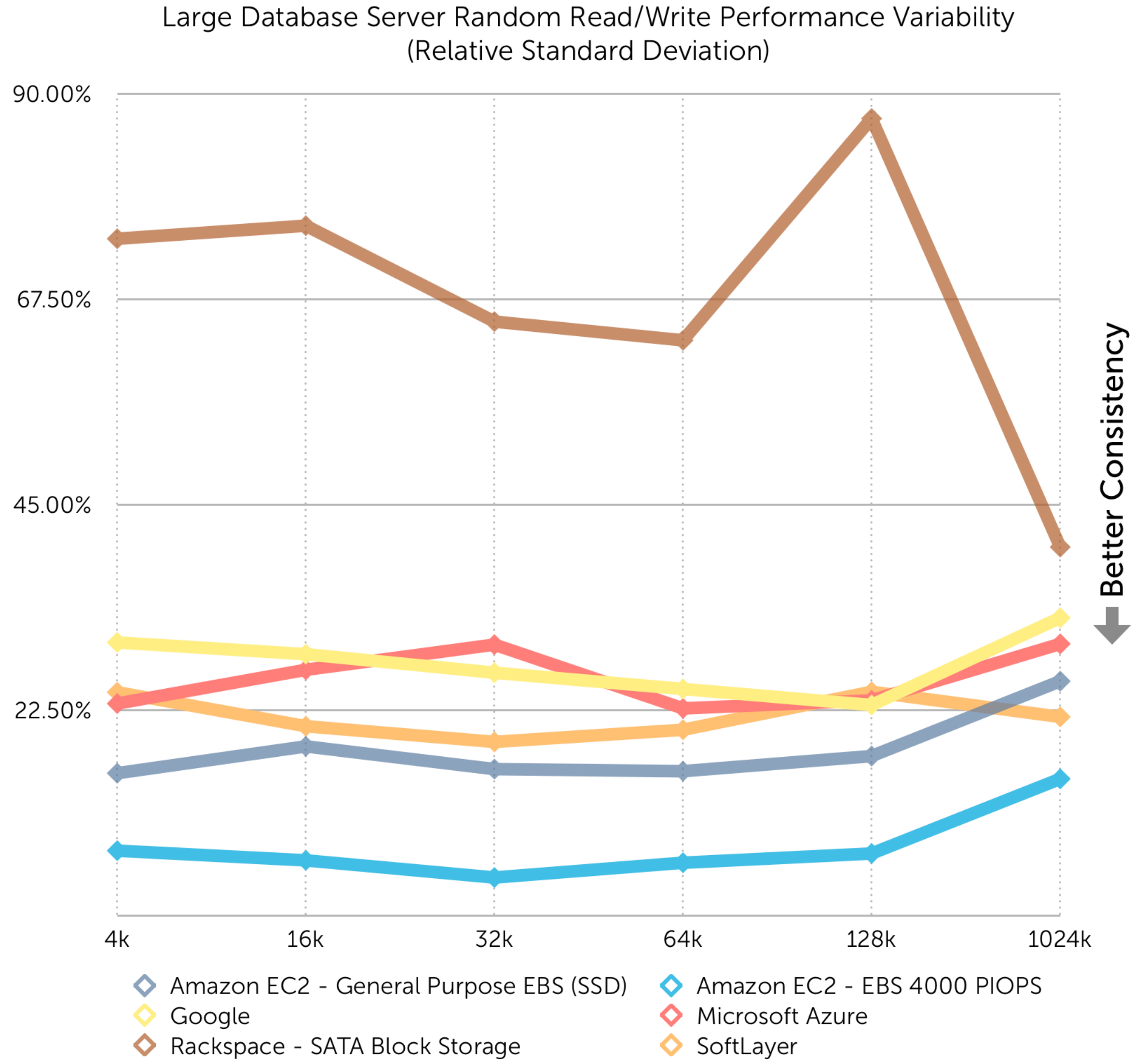
Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 4000 PIOPS (KB/s)	Google (KB/s)	Microsoft Azure	Rackspace - SATA Block Storage (KB/s)	SoftLayer (KB/s)
4k	1,173.67	16,485.00	276.67	NA	1,323.67	45,414.67
16k	4,595.00	62,624.00	1,071.00	NA	6,250.33	116,908.00
32k	8,587.33	65,284.67	2,045.33	NA	12,330.33	122,700.67
64k	14,195.33	65,287.33	3,758.00	NA	20,133.67	123,251.67
128k	14,264.00	59,710.67	6,436.33	NA	21,933.33	113,053.00
1024k	35,778.00	59,691.67	17,266.00	NA	55,184.00	107,788.00

Sequential Disk Read/Write Performance Consistency - Large Database Server



Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 4000 PIOPS	Google	Microsoft Azure	Rackspace - SATA Block Storage	SoftLayer
4k	13.47%	5.57%	23.00%	NA	75.09%	31.71%
16k	15.33%	6.84%	16.91%	NA	76.32%	17.72%
32k	25.28%	5.33%	16.88%	NA	63.92%	19.71%
64k	13.13%	4.94%	21.48%	NA	52.97%	21.73%
128k	13.83%	6.05%	18.49%	NA	47.86%	18.97%
1024k	24.79%	14.73%	29.15%	NA	40.12%	26.02%

Random Disk Read/Write Performance Consistency - Large Database Server



Block Size	Amazon EC2 - General Purpose EBS (SSD)	Amazon EC2 - EBS 4000 PIOPS	Google	Microsoft Azure	Rackspace - SATA Block Storage	SoftLayer
4k	15.58%	7.09%	29.91%	NA	74.14%	24.45%
16k	18.52%	6.03%	28.62%	NA	75.57%	20.73%
32k	16.03%	4.15%	26.59%	NA	65.06%	19.01%
64k	15.80%	5.77%	24.81%	NA	63.01%	20.34%
128k	17.46%	6.78%	23.05%	NA	87.31%	24.50%
1024k	25.66%	14.96%	32.61%	NA	40.36%	21.75%

Internal Network Performance

Our internal network testing measures latency and throughput between compute instances from the same provider in the same data center. The test results below list median latency (milliseconds - ms) and throughput (megabits per second - Mb/s) metrics for each provider and pairing of web and database servers: small, medium, and large. These metrics are based on single threaded tests. It might be possible to obtain higher throughput using multithreaded testing.

Internal Network Performance Comments

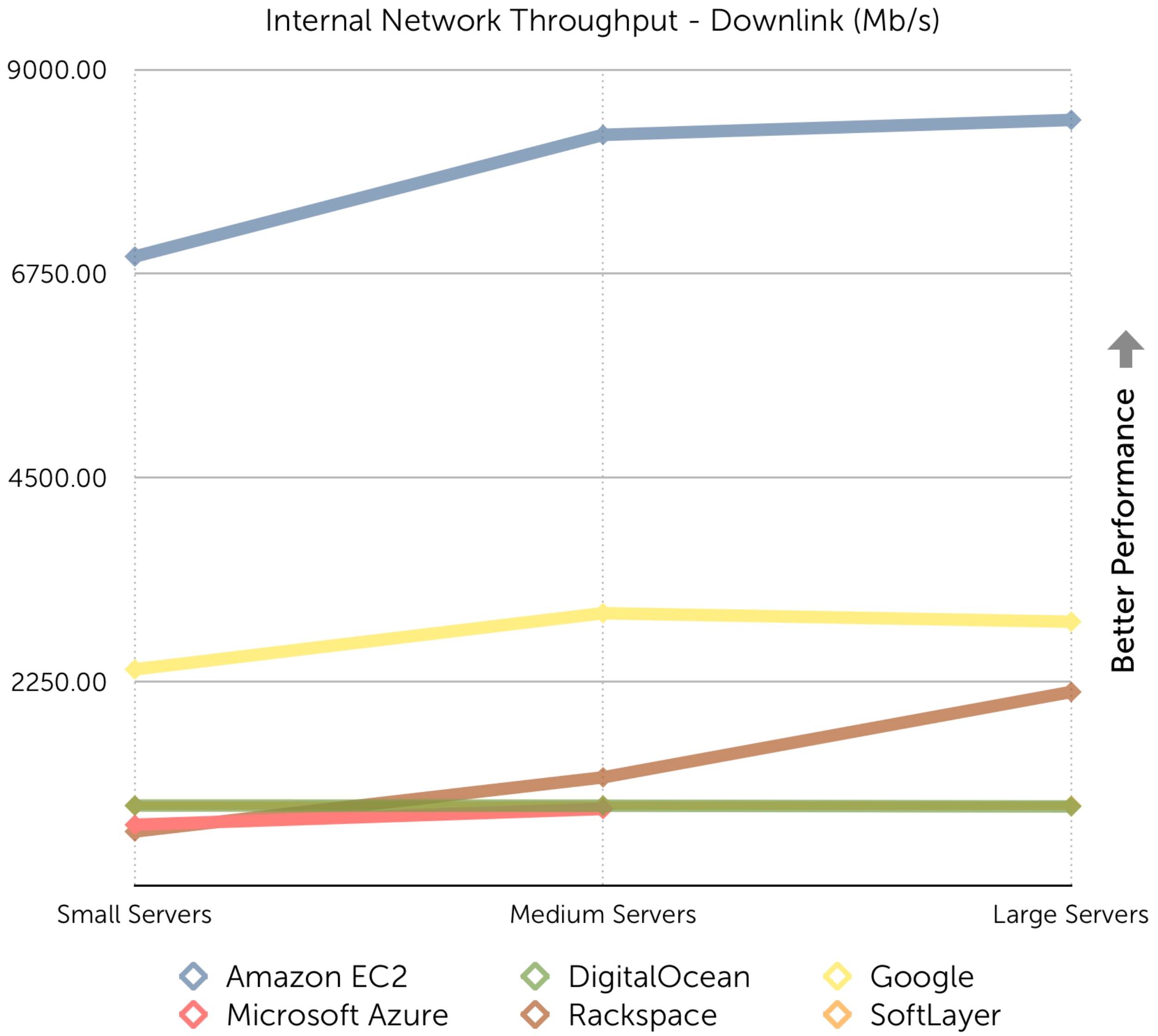
Optimal network settings that do not incur additional usage charges were chosen for each services. For Amazon EC2 this involved use of a placement group, virtual private network and instances with enhanced networking enabled. For Azure it involved use of an affinity group and virtual private network. For DigitalOcean it involved use of a region with support for private IP addresses (NY2). For Google it involved use of their most recent regional data center (Asia East) which according to their press release supports better networking performance. For SoftLayer it involved use of the fastest compute instance network option - one-gigabit private and public networks. Internal (private) networks and IP addresses were used for all network tests.

Based on our testing, DigitalOcean, SoftLayer, and Microsoft Azure internal networks appear limited to 1-gigabit per second (Gb/s), while Amazon EC2, Google, and Rackspace support higher throughput. As advertised, use of Amazon EC2 placement groups, virtual private networks and enhanced networking appear to provide nearly 10-gigabit throughput capacity and low latency. Amazon EC2 networking performance is slower without use of these optimal settings. If your workload calls for transferring large amounts of data between compute instances, this is worth considering.

Additionally, Amazon EC2, Google, Microsoft Azure and Rackspace appear to limit throughput based on size of compute instance, with large compute instances supporting higher throughput. In the comparisons of these services, performance is limited by the smaller of the two instances, which in our case is the web server. Throughput for DigitalOcean and SoftLayer was similar for all instance types. Rackspace instances also appear to have asymmetrical throughput characteristics where the instance downlink is capable of higher throughput than the uplink (which equates to the server’s download speed).

Internal Network Performance - Downlink

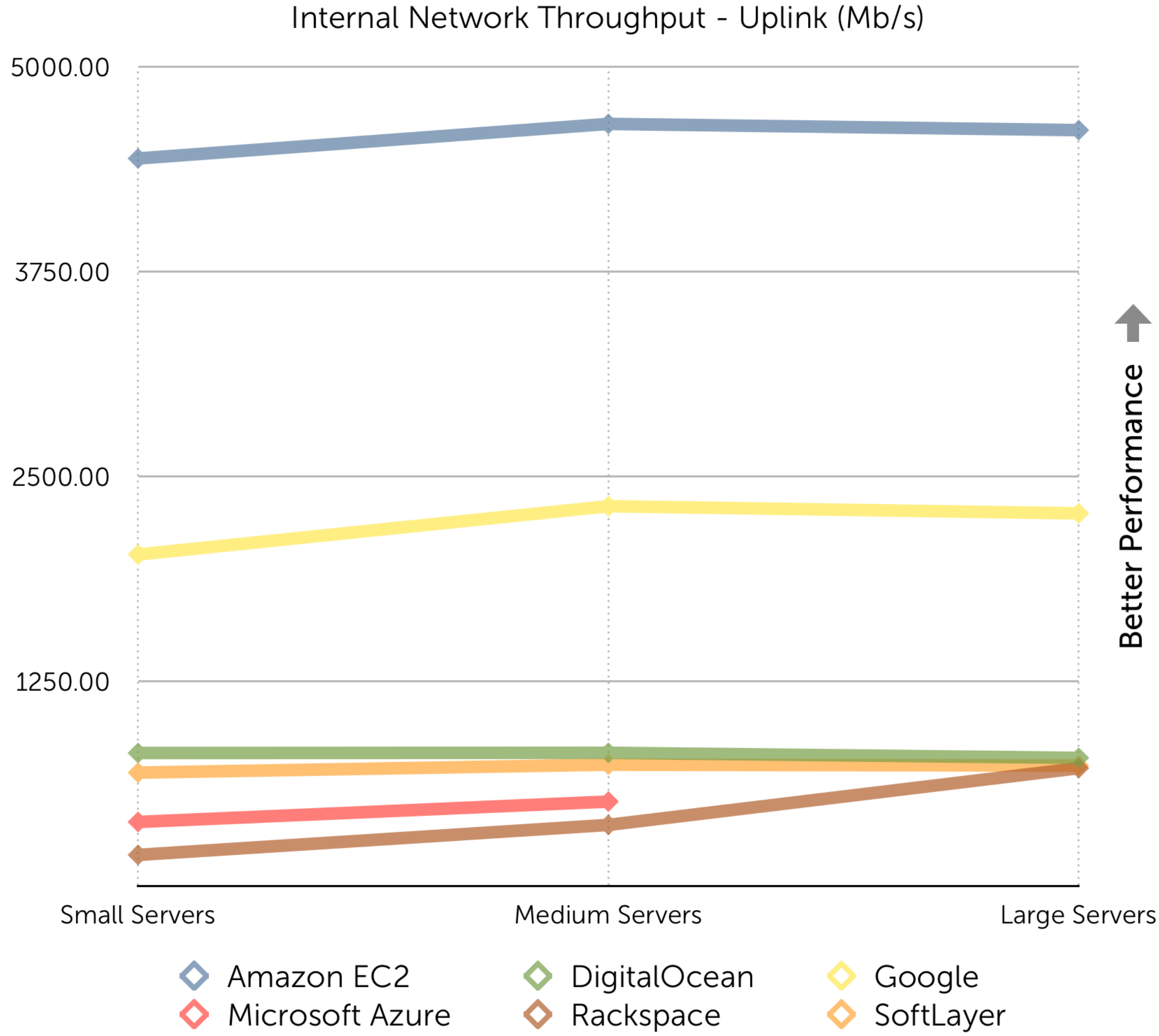
The analysis below lists median downlink throughput for each pairing of web and database servers: small, medium and large. Throughput is represented in megabits per second, where a higher value is better. Link direction is from the web server's perspective, meaning the metrics represent how fast the web server can transfer data from the database server.



Server Type	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Servers	6,939.35	883.34	2,385.47	671.28	597.13	882.52
Medium Servers	8,281.47	882.13	3,005.01	845.65	1,194.30	883.23
Large Servers	8,445.81	873.35	2,911.08	NA	2,136.75	881.60

Internal Network Performance - Uplink

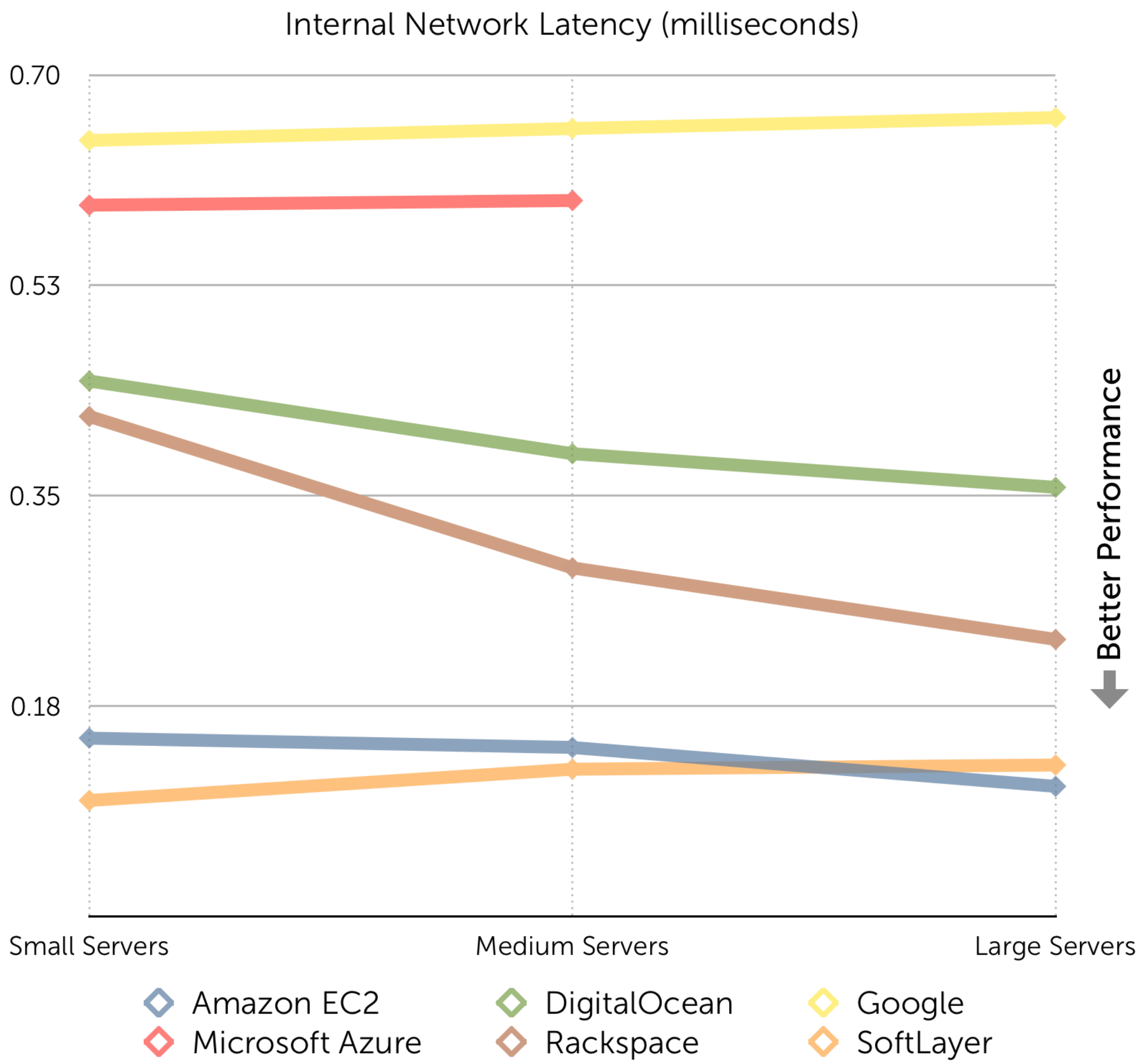
The analysis below lists median uplink throughput for each pairing of web and database servers: small, medium and large. Throughput is represented in megabits per second, where a higher value is better. Link direction is from the web server's perspective, meaning the metrics represent how fast the web server can transfer data to the database server.



Server Type	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Servers	4,440.39	810.30	2,022.23	389.32	188.10	690.63
Medium Servers	4,651.54	811.54	2,317.27	513.72	371.34	739.23
Large Servers	4,613.53	780.62	2,273.53	NA	717.66	733.34

Internal Network Performance - Latency

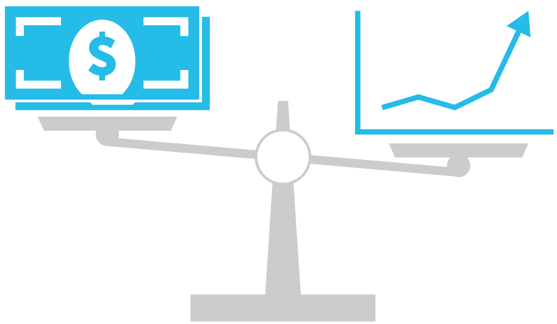
The analysis below lists median latency for each pairing of web and database servers: small, medium, and large. Latency is represented in milliseconds, where a lower value is better.



Server Type	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
Small Servers	0.15	0.45	0.65	0.59	0.42	0.10
Medium Servers	0.14	0.39	0.66	0.60	0.29	0.12
Large Servers	0.11	0.36	0.66	NA	0.23	0.13

Value Comparisons

In this section we present an approach to choosing a cloud service on the basis of value, where value is a ratio of cost to relevant performance characteristics.



Cost is often a significant decision-making factor. If it were not, you would probably choose the largest and fastest compute instance available for even small workloads. But in choosing a provider, you should look for a compute instance that achieves your performance goals in a cost-effective manner. In other words, you look for a compute instance that offers good value.

- What is the actual cost of compute instances?
- Which performance characteristics are relevant to my workload, and to what degree?
- What are my performance goals; that is, what is good enough?

To answer the first question, you calculate a single or normalized cost for each of the compute instances that you are comparing. Take into account the different pricing models used by each provider, including commit-based discounting and additional fees.

The second question we addressed in the "Benchmark Selection" section. The trick now is to reduce all the metrics produced by those benchmarks to a single value that you can use for comparison, which we discuss later in this report.

The last question is more arbitrary, but you must determine a cutoff for performance below which you will not consider compute instances. Using the performance metric from question two might help you to establish this threshold.

To quantify value, we can create a ratio of cost to performance by using the values derived from the first two questions. Then we can use this ratio to determine which provider and compute instances offer the best value. In this section, we'll step through this approach using the workloads and compute instances we've already selected.

Cost Normalization

Pricing per hour (or sometimes per minute) is the most common pricing model for compute services. This is called on-demand or utility pricing, because you pay only for what you actually use. (Some providers might offer additional pricing models if you commit to a set monthly spend, or to using compute instances for a one- or three-year term.) Your goal is to choose the pricing model that provides you with the best economic benefit, while still providing flexibility. It's also important to avoid getting locked in to spending for resources that you may or may not use.

ANNUAL
+ PRE-PAY
+ VOLUME DISCOUNT
+ ADD-ONS

$/HR

The method we suggest for comparing the costs of compute services with multiple pricing models is to convert non-hourly pricing to hourly pricing. Then you can use the resulting "normalized hourly value" to compare different price models and compute services.

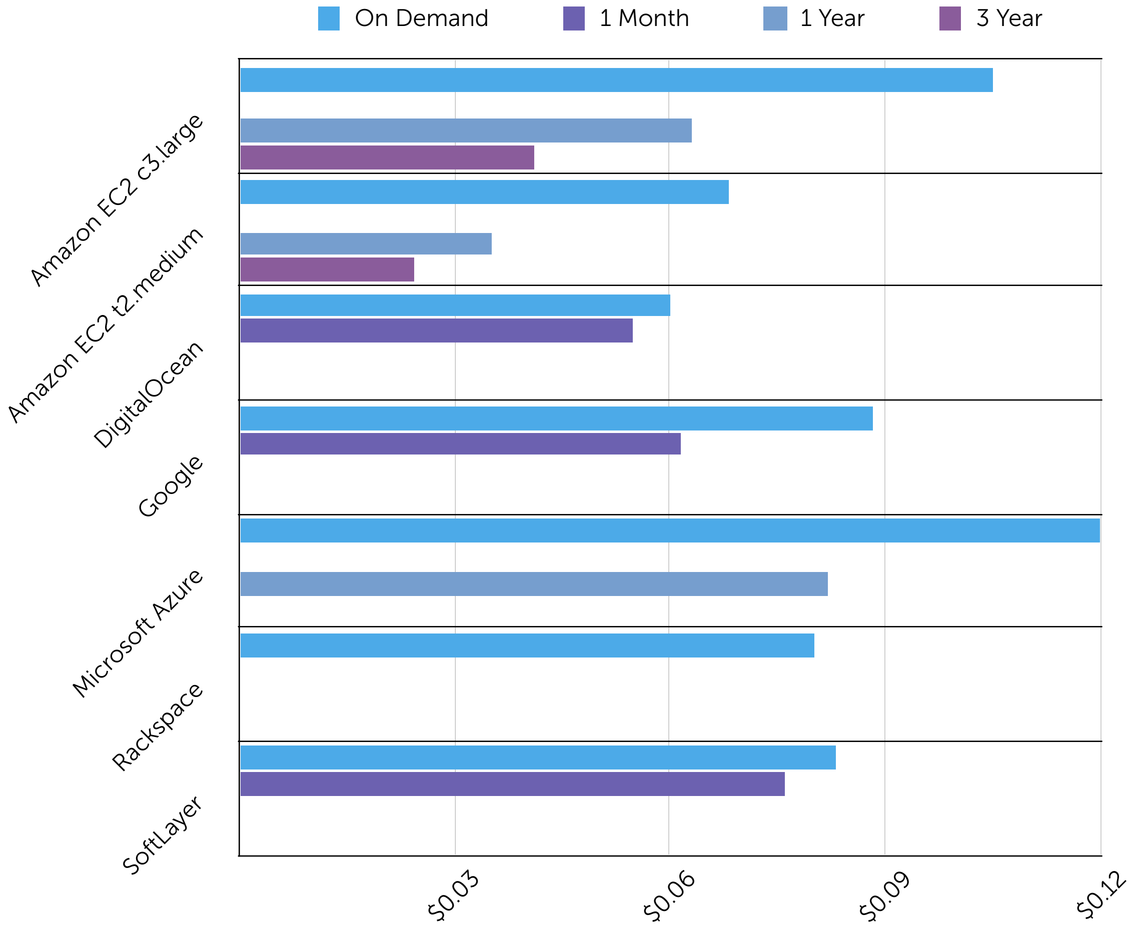
The following tables list a normalized hourly cost for each provider and compute instance based on all pricing models offered. Only discounts applicable to our selection of servers are included. Most providers also offer high-volume discounts. For example, Rackspace offers a 7-33 percent discount starting at $5,000 monthly, and AWS offers a 10-20 percent starting at $250,000 monthly. Volume discounts are not included in the following tables, because our combinations of servers would not reach the required monthly spend. A few more notes regarding pricing:

- We calculated prices based on the lowest cost US data center and a free operating system, such as CentOS.
- We list only published commit prices. Undisclosed or negotiated pricing may be possible with some providers.
- We used 730 hours per month and 8,760 hours per year as the basis for commit calculations.
- AWS offers three reserve-price usage models: light, medium, and heavy. We based prices in the commit columns on the heavy (full-time) usage model.
- We based the Microsoft Azure one-year commit pricing on the pre-paid option with a minimum $500 monthly spend.
- We based Google Compute Engine pricing on full-month sustained-use discounting.

Web Server Normalized Cost Comparisons

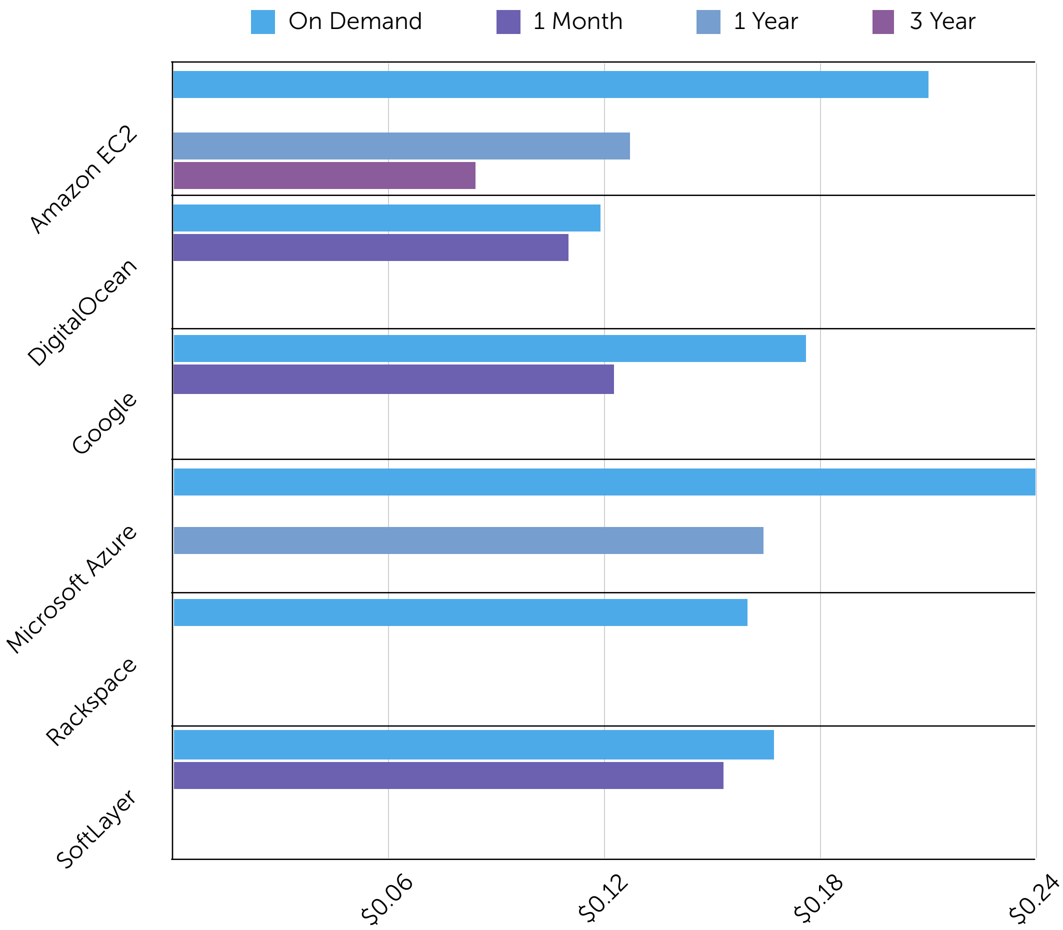
The proceeding graphs and tables list normalized hourly costs for each service and web server compute instance based on four pricing models. If a service does not support a pricing model, that column is left blank. Pricing is based on the lowest regional pricing available from each provider.

Small Web Server Normalized Cost



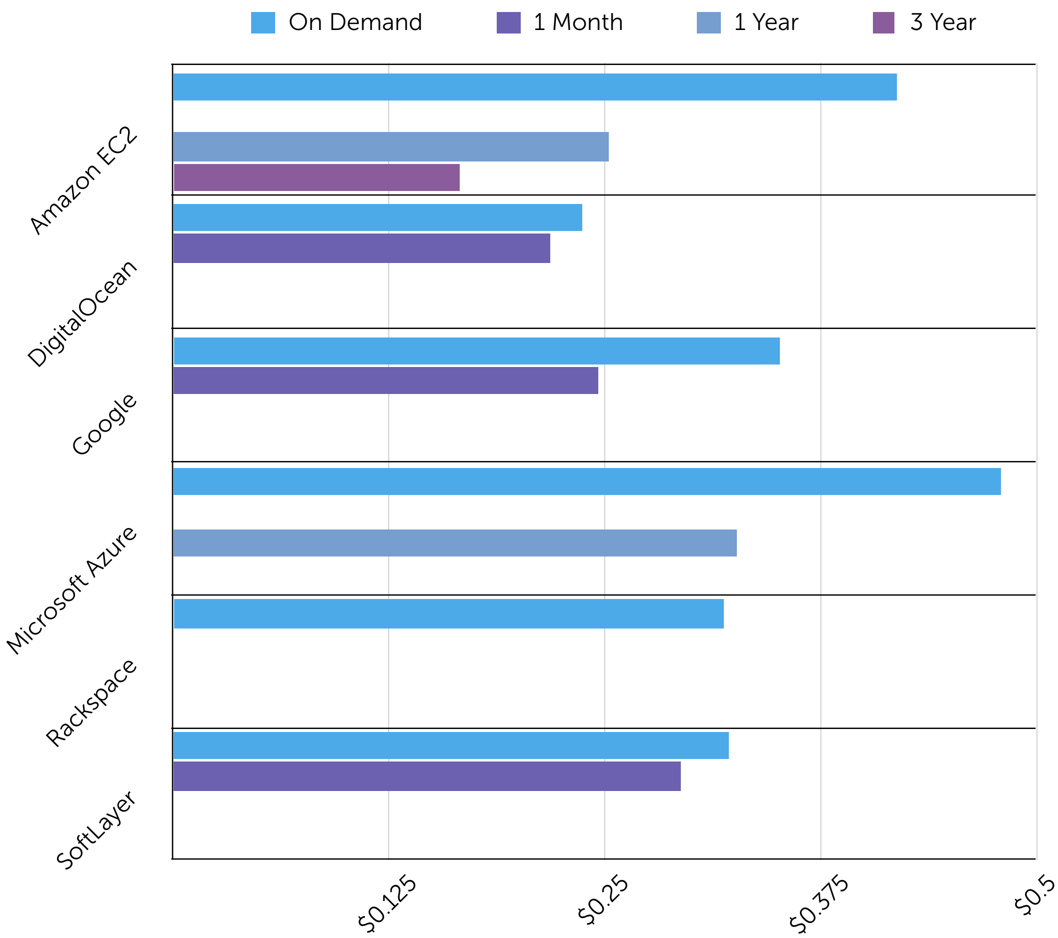
Service	Instance Type	On Demand	1 Month	1 Year	3 Year
Amazon EC2	c3.large	$0.105	NA	$0.063	$0.041
Amazon EC2	t2.medium	$0.0680	NA	$0.0353	$0.0246
DigitalOcean	4 GB / 2 Cores	$0.06	$0.055	NA	NA
Google	n1-highcpu-2	$0.088	$0.0616	NA	NA
Microsoft Azure	Medium (A2)	$0.12	NA	$0.08	NA
Rackspace	Performance 1 2GB	$0.08	NA	NA	NA
SoftLayer	2 GB / 2 Cores	$0.083	$0.076	NA	NA

Medium Web Server Normalized Cost



Service	Instance Type	On Demand	1 Month	1 Year	3 Year
Amazon EC2	c3.xlarge	$0.21	NA	$0.127	$0.084
DigitalOcean	8 GB / 4 Cores	$0.119	$0.11	NA	NA
Google	n1-highcpu-4	$0.176	$0.1232	NA	NA
Microsoft Azure	Large (A3)	$0.24	NA	$0.16	NA
Rackspace	Performance 1 4GB	$0.16	NA	NA	NA
SoftLayer	4 GB / 4 Cores	$0.167	$0.153	NA	NA

Large Web Server Normalized Cost

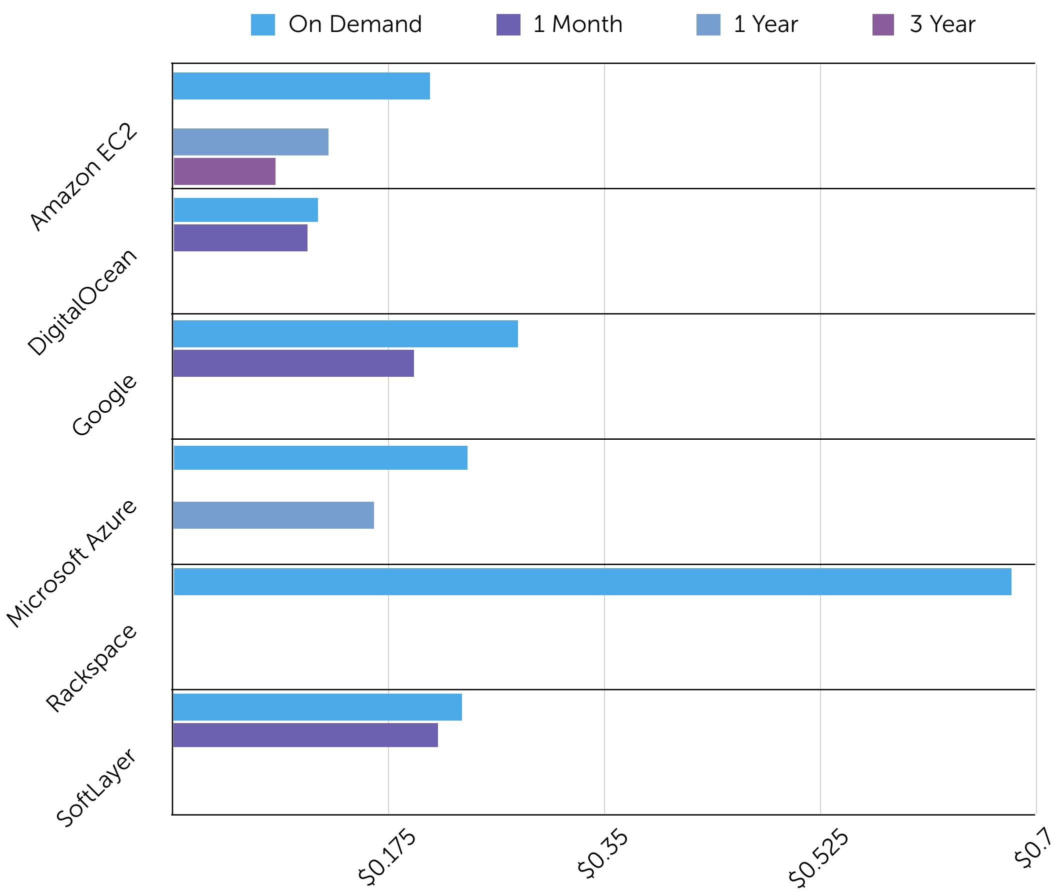


Service	Instance Type	On Demand	1 Month	1 Year	3 Year
Amazon EC2	c3.2xlarge	$0.42	NA	$0.253	$0.167
DigitalOcean	16 GB / 8 Cores	$0.238	$0.219	NA	NA
Google	n1-highcpu-8	$0.352	$0.2464	NA	NA
Microsoft Azure	Extra Large (A4)	$0.48	NA	$0.33	NA
Rackspace	Performance 1 8GB	$0.32	NA	NA	NA
SoftLayer	8 GB / 8 Cores	$0.323	$0.295	NA	NA

Database Server Normalized Cost Comparisons

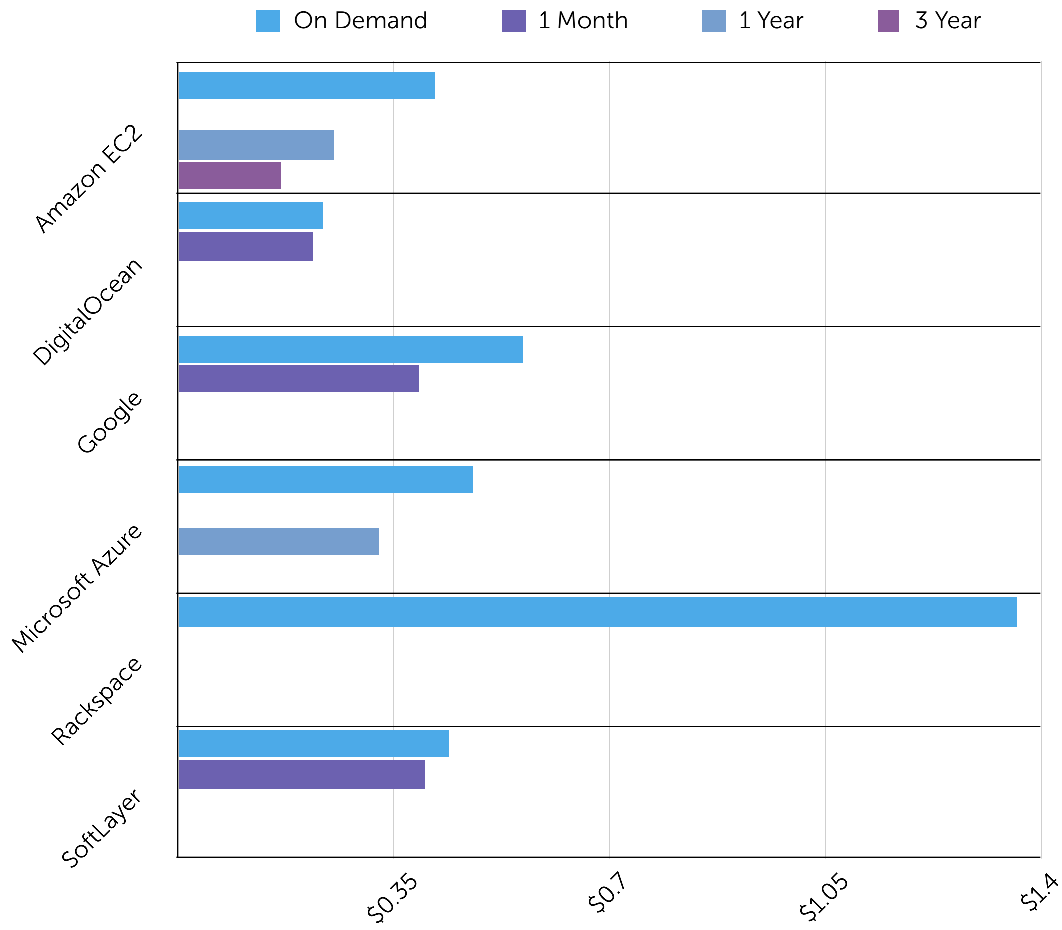
The proceeding graphs and tables list normalized hourly costs for each service and database server compute instance based on four pricing models. If a service does not support a pricing model, that column is left blank. Pricing is based on the lowest regional pricing available from each provider.

Small Database Server Normalized Cost



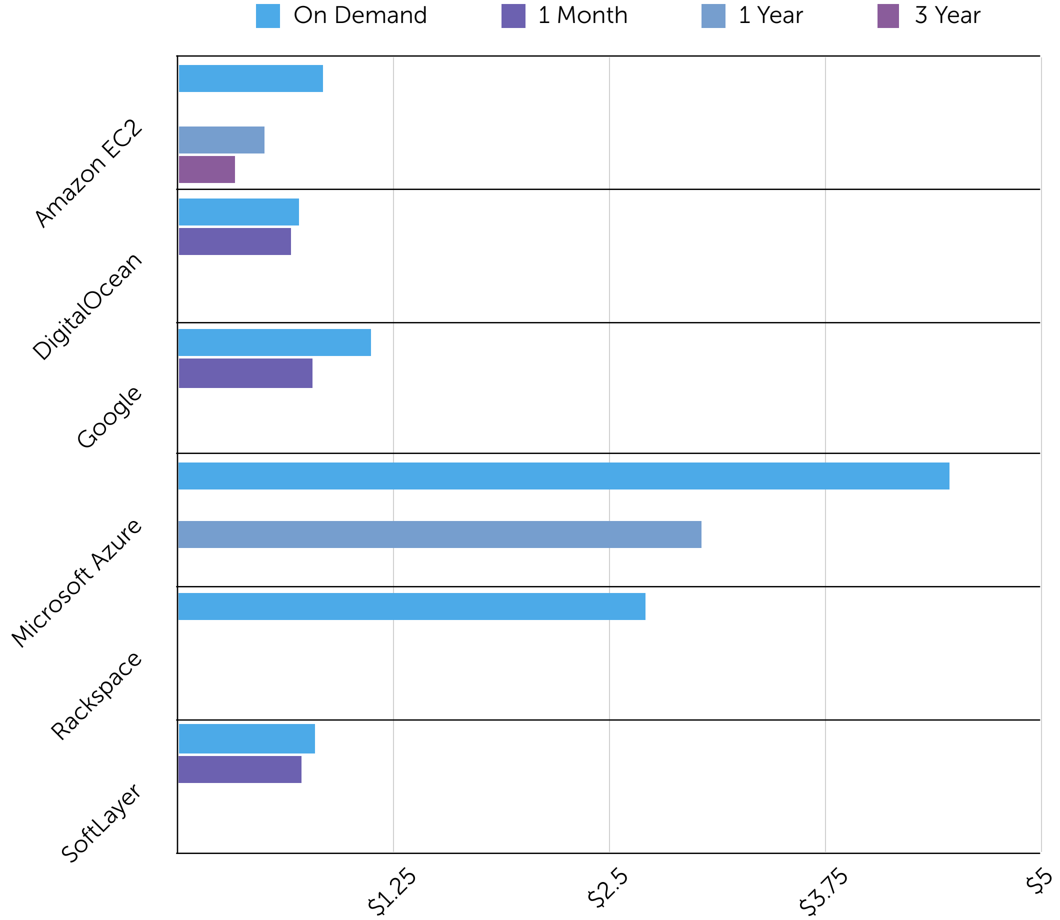
Service	Instance Type	On Demand	1 Month	1 Year	3 Year
Amazon EC2	c3.xlarge	$0.21	NA	$0.127	$0.084
DigitalOcean	8 GB / 4 Cores	$0.119	$0.11	NA	NA
Google	n1-standard-4	$0.28	$0.196	NA	NA
Microsoft Azure	Large (A3)	$0.24	NA	$0.16	NA
Rackspace	Performance 2 15GB	$0.68	NA	NA	NA
SoftLayer	8 GB / 4 Cores	$0.236	$0.216	NA	NA

Medium Database Server Normalized Cost



Service	Instance Type	On Demand	1 Month	1 Year	3 Year
Amazon EC2	c3.2xlarge	$0.42	NA	$0.253	$0.167
DigitalOcean	16 GB / 8 Cores	$0.238	$0.219	NA	NA
Google	n1-standard-8	$0.56	$0.392	NA	NA
Microsoft Azure	Extra Large (A4)	$0.48	NA	$0.33	NA
Rackspace	Performance 2 30GB	$1.36	NA	NA	NA
SoftLayer	16 GB / 8 Cores	$0.438	$0.40	NA	NA

Large Database Server Normalized Cost



Service	Instance Type	On Demand	1 Month	1 Year	3 Year
Amazon EC2	c3.4xlarge	$0.84	NA	$0.507	$0.335
Microsoft Azure	A9	$4.47	NA	$3.04	NA
DigitalOcean	48 GB / 16 Cores	$0.705	$0.658	NA	NA
Google	n1-standard-16	$1.12	$0.784	NA	NA
Rackspace	Performance 2 60GB	$2.72	NA	NA	NA
SoftLayer	32 GB / 16 Cores	$0.794	$0.726	NA	NA

Database Server Storage Costs

The proceeding tables list storage costs for each service and database server instance. Pricing is based on the lowest regional pricing available from each provider.

Small Database Server Storage Cost

Service	Storage Type	Hourly Cost
Amazon EC2	50 GB General Purpose EBS (SSD) 50 GB EBS Optimized w/ 1,500 PIOPS	\$0.003 \$0.234
DigitalOcean	Not Available	NA
Google	50 GB Persistent Disk	\$0.003
Microsoft Azure	50 GB Locally Redundant	\$0.005
Rackspace	50 GB Standard Cloud Block Storage 50 GB SSD Cloud Block Storage	\$0.008 \$0.034
SoftLayer	250 GB Portable SAN Volume	\$0.05

Medium Database Server Storage Cost

Service	Storage Type	Hourly Cost
Amazon EC2	100 GB General Purpose EBS (SSD) 100 GB EBS Optimized w/ 3,000 PIOPS	\$0.007 \$0.428
DigitalOcean	Not Available	NA
Google	100 GB Persistent Disk	\$0.006
Microsoft Azure	100 GB Locally Redundant	\$0.01
Rackspace	100 GB Standard Cloud Block Storage 100 GB SSD Cloud Block Storage	\$0.016 \$0.069
SoftLayer	250 GB Portable SAN Volume	\$0.05

Large Database Server Storage Cost

Service	Storage Type	Hourly Cost
Amazon EC2	200 GB General Purpose EBS (SSD) 200 GB EBS Optimized w/ 4,000 PIOPS	$0.014 $0.582
DigitalOcean	Not Available	NA
Google	200 GB Persistent Disk	$0.012
Microsoft Azure	200 GB Locally Redundant	$0.02
Rackspace	200 GB Standard Cloud Block Storage 200 GB SSD Cloud Block Storage	$0.032 $0.137
SoftLayer	250 GB Portable SAN Volume	$0.05

Performance Normalization

To calculate value, we must first reduce performance to a single numeric value. We can then use this metric and the preceding costs to numerically quantify value for each compute instance and pricing model.

Because we've compared performance using multiple benchmarks, and each benchmark metric is based on a different numeric scale, before combining them we must first convert them a common scale. To make this conversion we use a reference system as a normalizing constant. By dividing the benchmark metrics of compute instances by the same metrics of the reference system we produce ratios that represent the performance of each compute instance relative to the reference system.

The reference system we use in the analysis below is a bare metal (non-virtualized) Dell M610 PowerEdge server. This server was configured with two Intel X5650 2.66 GHz processors, 48 GB DDR3-10166 memory, and a Seagate SAS 2.0 10k RPM drive. We ran the same benchmarks on this server and used the metrics to produce the performance ratios. We multiplied ratios by 100, meaning that a value of exactly 100 signifies performance that is nearly equal to the reference system, less than 100 worse, and greater than 100 better.

The proceeding analysis presents performance and value with respect to CPU. Both SPECint and SPECfp Rate/Base 2006 benchmark metrics from the SPEC CPU 2006 benchmark suite (29 total benchmarks) are used to produce the single reduced performance metrics. These metrics are then divided by the normalized costs in order to quantify the value of each compute instance where a higher value metric represents better value. Value metrics are provided for each compute instance and all four pricing models (on-demand, 1-month, 1-year, and 3-year commit). If a service does not support one of the pricing models, the preceding model is used.

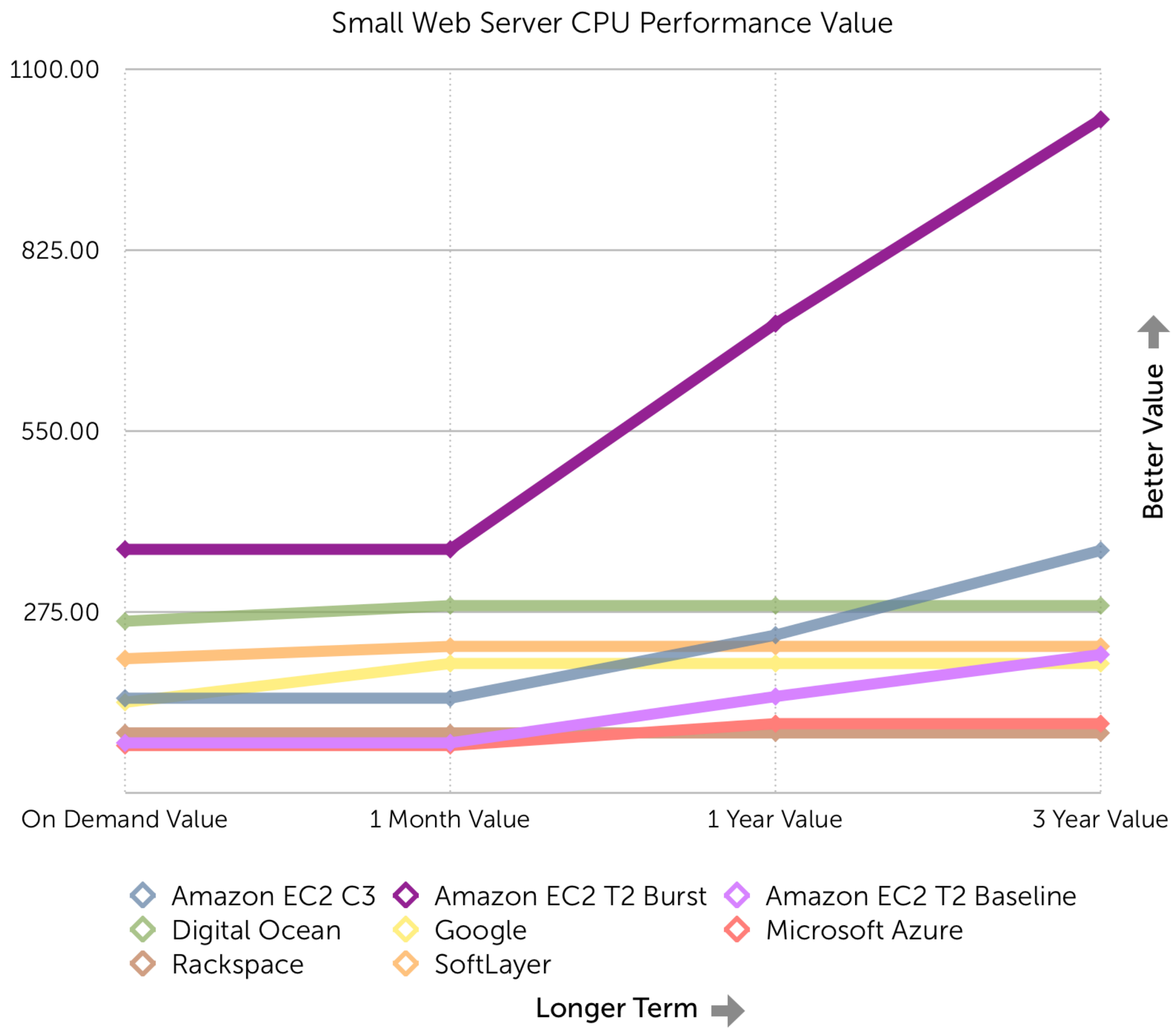
We chose to focus on CPU performance to quantify value because this is generally the way that providers price compute services. This analysis does not factor in other performance criteria such as disk and network, nor differences in memory configuration. It's designed simply as a starting point for evaluating the services.

Web Server Value Comparisons

The following tables list the results of our small, medium, and large database server value comparisons. Keep in mind that a higher value indicates better performance.

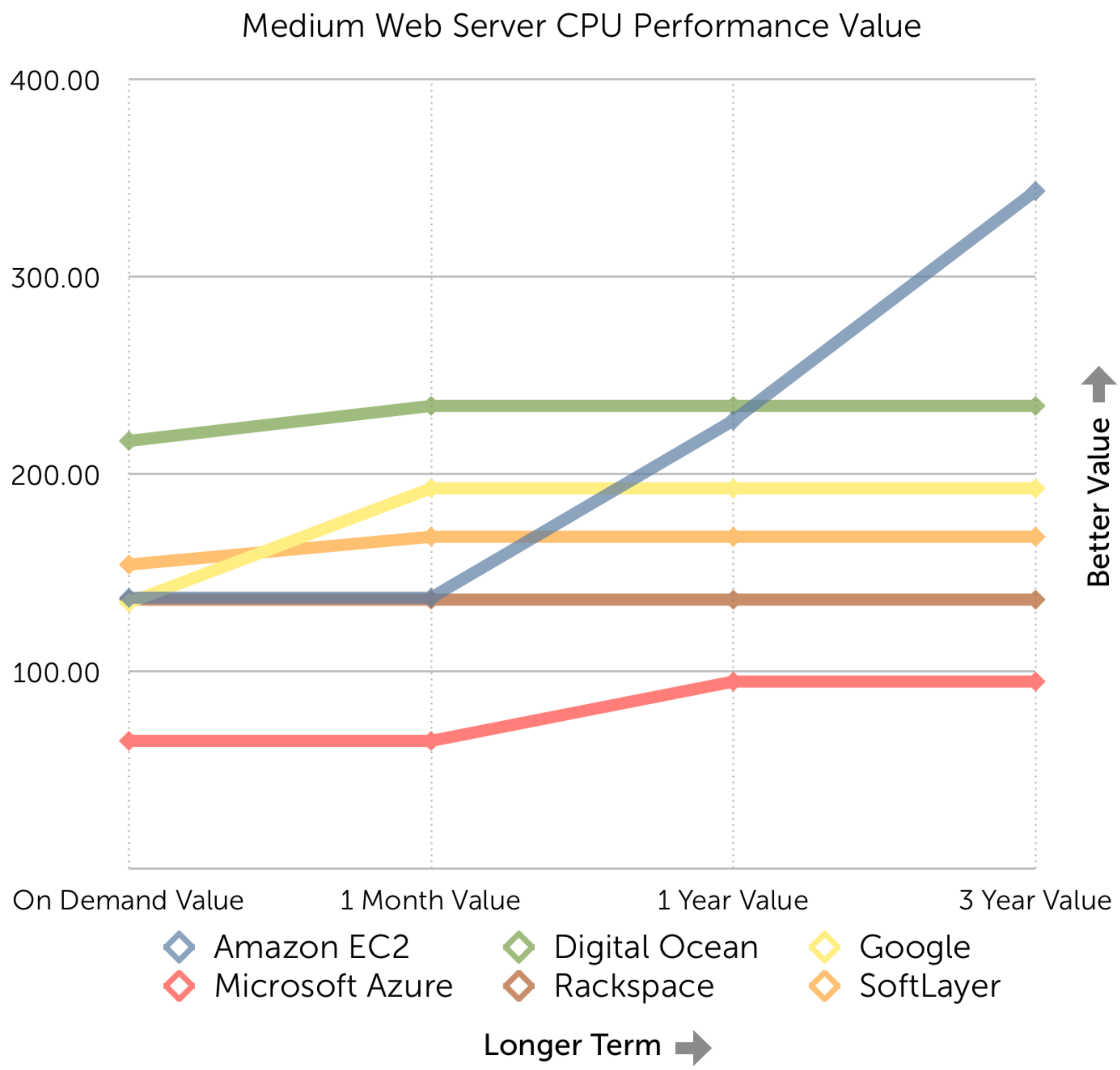
Small Web Server Value - Graph

This graph includes value analysis for two Amazon EC2 instances - c3.large and t2.medium. The t2.medium value analysis includes separate measurements for peak and baseline operating modes. t2.medium peak operating mode is supports better CPU performance for approximately 20% of the total operating time.



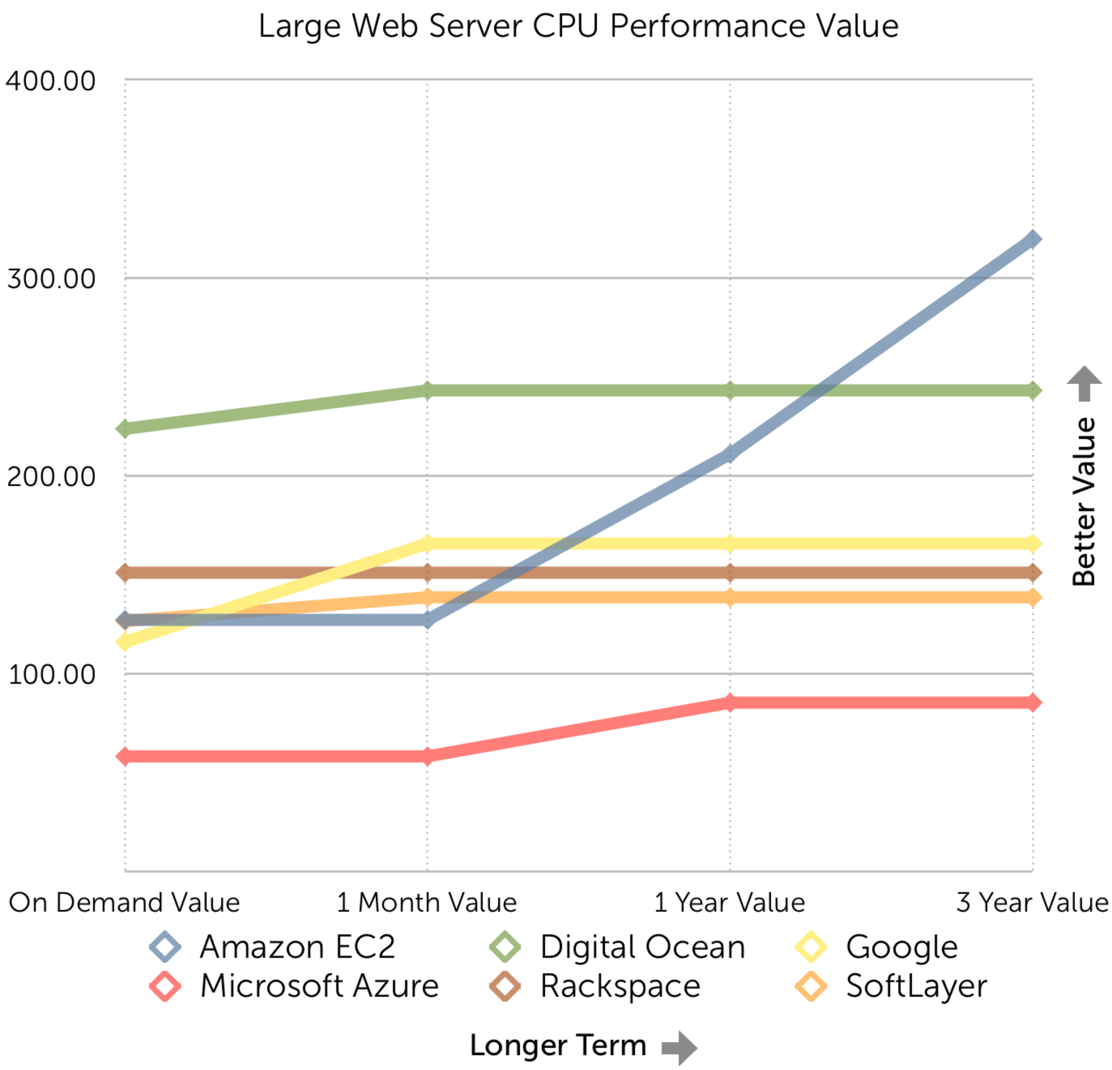
Service	CPU Performance Ratio	On Demand Value	1 Month Value	1 Year Value	3 Year Value
Amazon EC2 C3	15.11	143.90	143.90	239.84	368.53
Amazon EC2 T2 Burst	25.17	370.15	370.15	713.28	1023.56
Amazon EC2 T2 Baseline	5.17	76.03	76.03	146.51	210.24
Digital Ocean	15.65	260.92	284.64	284.64	284.64
Google	12.12	137.67	196.68	196.68	196.68
Microsoft Azure	8.63	71.95	71.95	105.29	105.29
Rackspace	7.29	91.11	91.11	91.11	91.11
SoftLayer	16.92	203.91	222.69	222.69	222.69

Medium Web Server Value



Service	CPU Performance Ratio	On Demand Value	1 Month Value	1 Year Value	3 Year Value
Amazon EC2	28.85	137.40	137.40	227.19	343.49
Digital Ocean	25.81	216.86	234.61	234.61	234.61
Google	23.75	134.97	192.81	192.81	192.81
Microsoft Azure	15.57	64.87	64.87	94.94	94.94
Rackspace	21.83	136.44	136.44	136.44	136.44
SoftLayer	25.75	154.19	168.30	168.30	168.30

Large Web Server Value

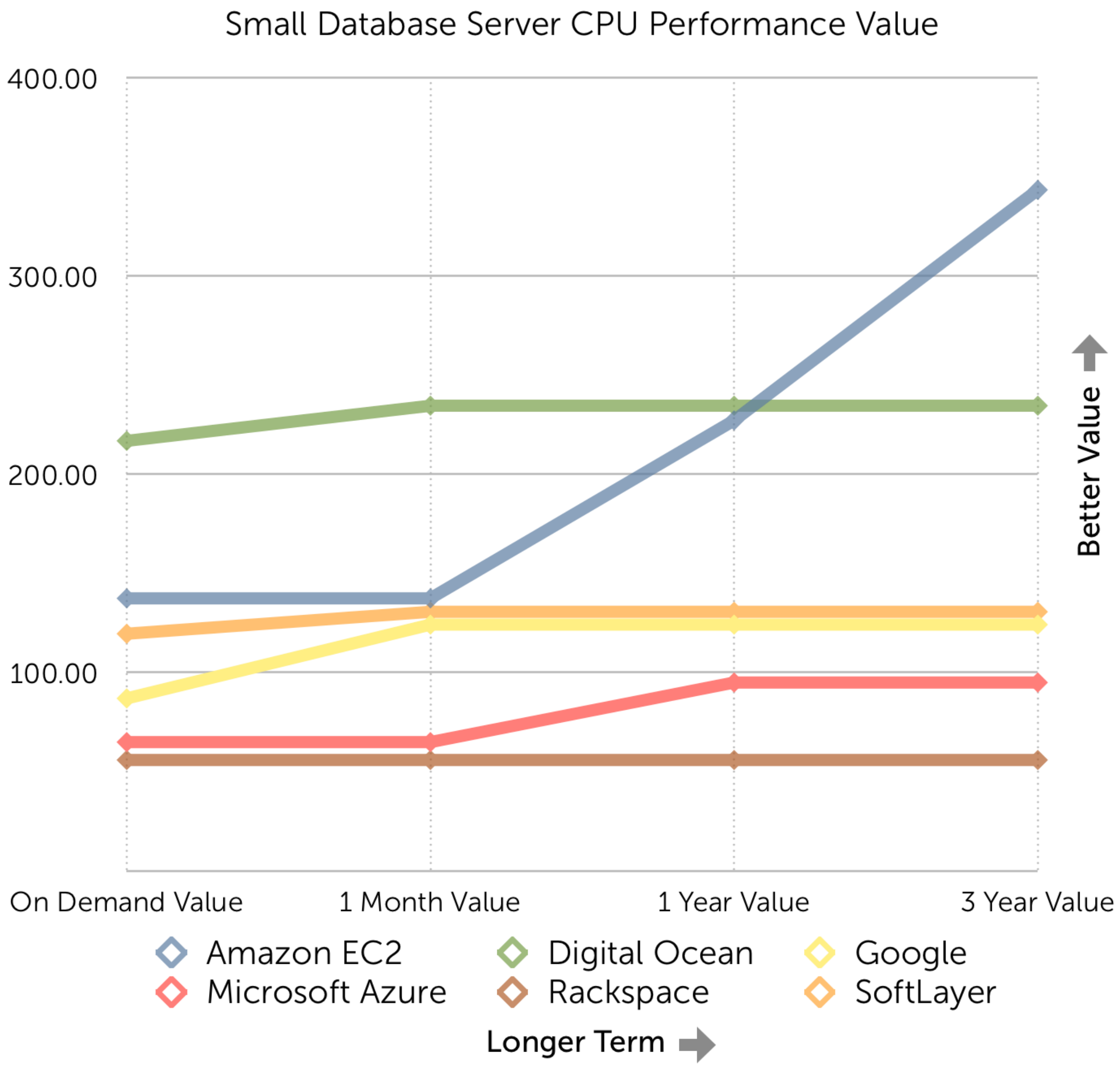


Service	CPU Performance Ratio	On Demand Value	1 Month Value	1 Year Value	3 Year Value
Amazon EC2	53.35	127.03	127.03	210.89	319.49
Digital Ocean	53.22	223.60	243.00	243.00	243.00
Google	40.81	115.92	165.61	165.61	165.61
Microsoft Azure	27.86	58.05	58.05	85.21	85.21
Rackspace	48.28	150.88	150.88	150.88	150.88
SoftLayer	40.84	126.44	138.44	138.44	138.44

Database Server Value Comparisons

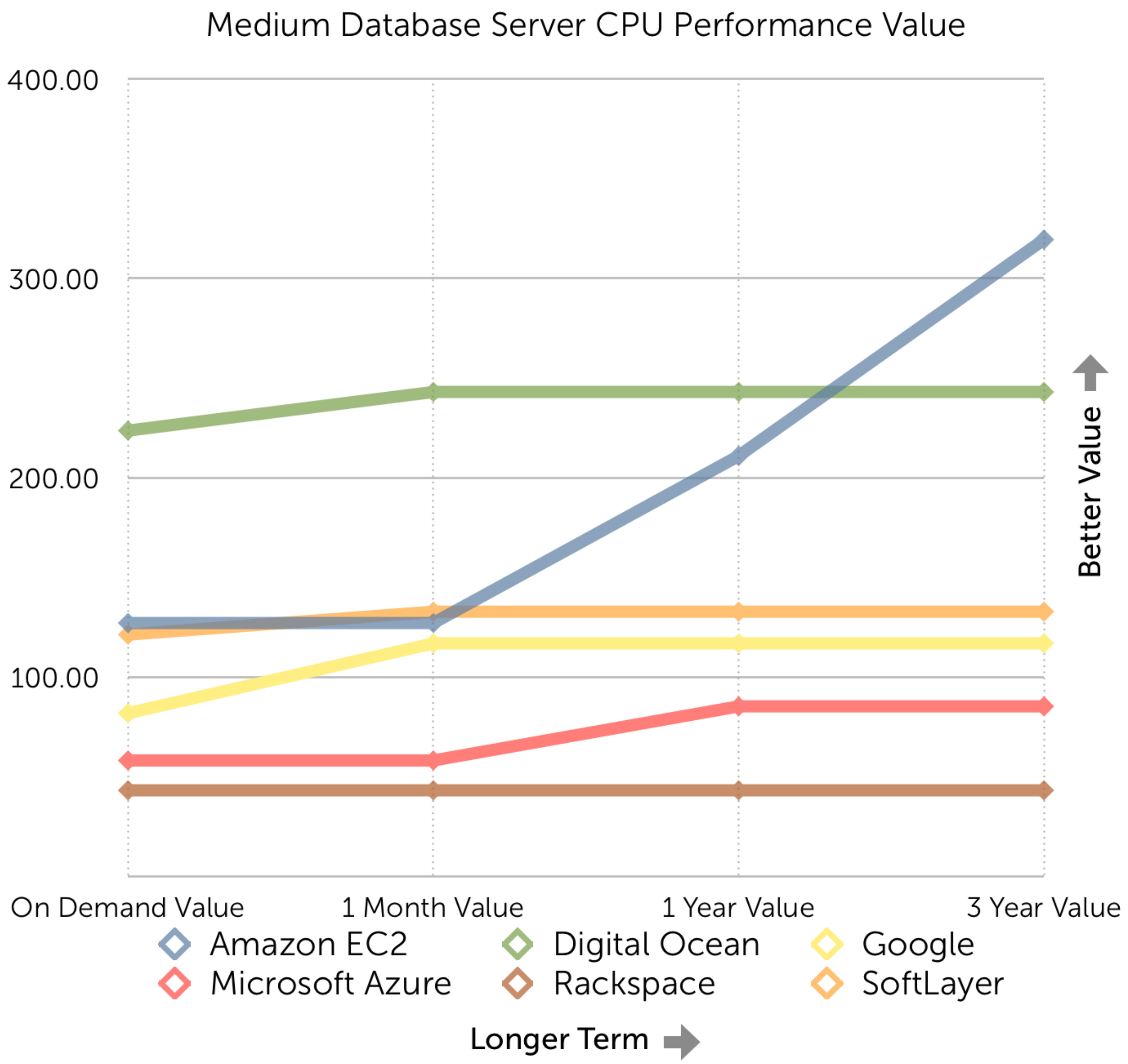
The following tables list the results of our small, medium, and large database server value comparisons. Keep in mind that a higher value indicates better performance.

Small Database Server Value



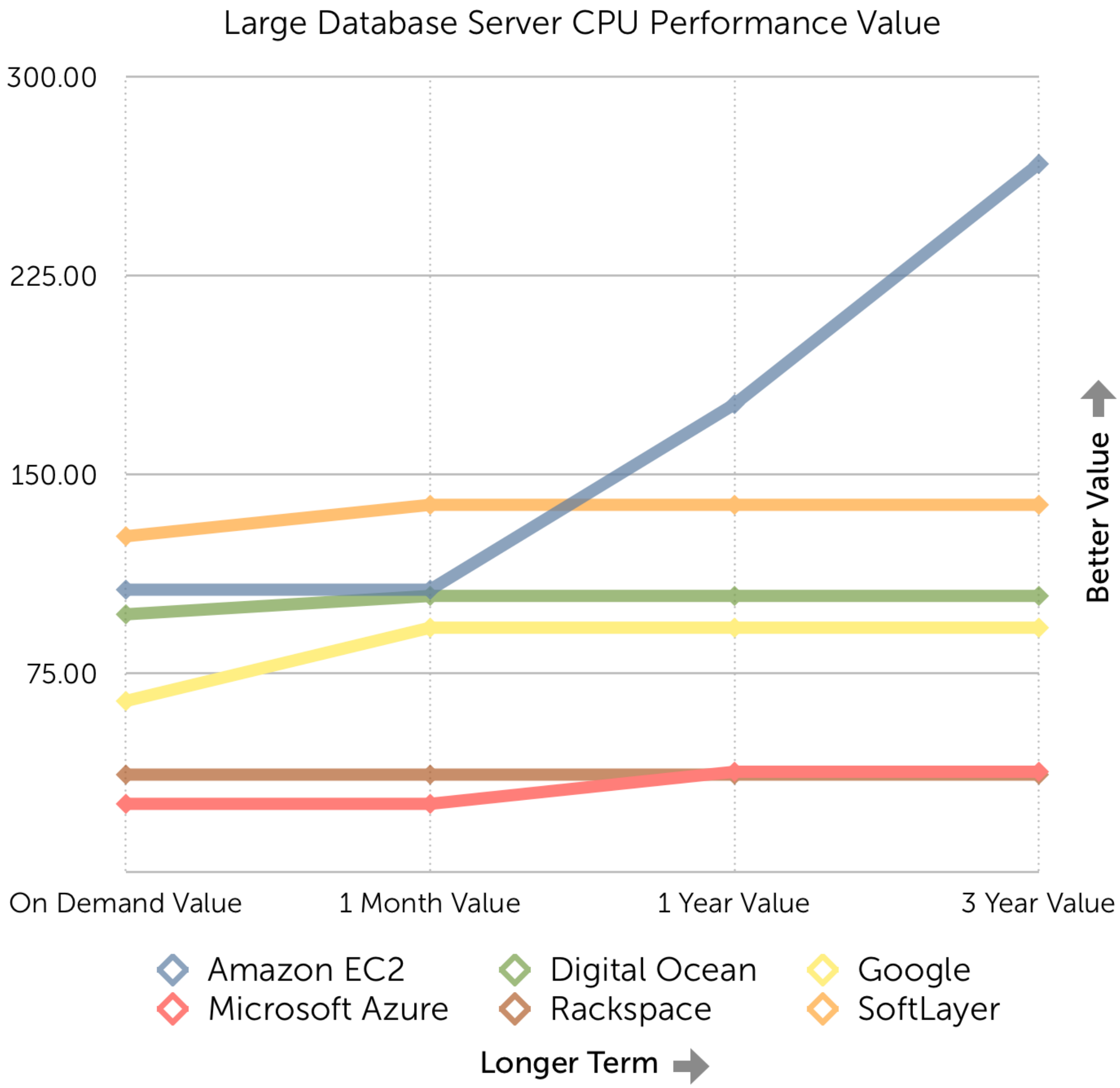
Service	CPU Performance Ratio	On Demand Value	1 Month Value	1 Year Value	3 Year Value
Amazon EC2	28.85	137.40	137.40	227.19	343.49
Digital Ocean	25.81	216.86	234.61	234.61	234.61
Google	24.34	86.92	124.17	124.17	124.17
Microsoft Azure	15.57	64.87	64.87	94.94	94.94
Rackspace	39.44	58.01	58.01	58.01	58.01
SoftLayer	28.21	119.55	130.62	130.62	130.62

Medium Database Server Value



Service	CPU Performance Ratio	On Demand Value	1 Month Value	1 Year Value	3 Year Value
Amazon EC2	54.39	129.50	129.50	214.98	325.68
Digital Ocean	48.54	203.95	221.65	221.65	221.65
Google	46.79	83.55	119.36	119.36	119.36
Microsoft Azure	27.86	58.05	58.05	85.21	85.21
Rackspace	64.82	47.66	47.66	47.66	47.66
SoftLayer	52.55	119.98	131.38	131.38	131.38

Large Database Server Value



Service	CPU Performance Ratio	On Demand Value	1 Month Value	1 Year Value	3 Year Value
Amazon EC2	89.48	106.52	106.52	176.48	267.10
Digital Ocean	68.54	97.22	104.17	104.17	104.17
Google	72.27	64.53	92.18	92.18	92.18
Microsoft Azure	115.06	25.74	25.74	37.85	37.85
Rackspace	100.01	36.77	36.77	36.77	36.77
SoftLayer	100.57	126.66	138.52	138.52	138.52

Other Considerations

When choosing a cloud provider, you must consider many factors in addition to performance and price. Cloud is an emerging technology, and although there are many commonalities between services, each offers a unique mix of features and capabilities.

Most providers start with a commercial or open source virtualization platform like Xen, OpenStack, or VMware, and build their own unique features on top of the platform. Some providers innovate aggressively, adding many additional features, while others have more basic virtualization capabilities. These additional features and capabilities are how many providers set themselves apart. This section reviews these features, capabilities, and the claims companies make about them.

Although new features are often very useful and cool, you should be cautious about adopting them unless you are 100 percent confident in your provider choice, because if only one provider offers a feature, that feature can lock you into using that provider.

Compute Instance Quotas

Compute services usually have default account quotas that limit the number of compute instances you can have. These quotas are usually based either on number of compute instances, total CPU cores, or memory.

Most providers allow you to request an increase of the default quota with some (typically larger providers) making this easier and quicker to obtain than others. Our experience with requesting such increases has been mixed including positive responses within hours from Amazon and Google, and region dependent infrastructure availability for SoftLayer.

Default quotas might give you a little insight into a provider's scale. A provider operating at a larger scale might offer more generous quotas. The following tables list the policies and quotas for each compute service. Quotas lists are for multiple given compute instance sizes within varying CPU cores and memory. Amazon EC2 and Google quotas are region-specific, so their columns show both total and regional limits (Amazon EC2 has 8 regions and Google 3).

Compute Service Quota Policies

Service	Policy
Amazon EC2	20 instances per region for most compute instance types
DigitalOcean	10 compute instances (Droplets)
Google	24 CPU cores per region
Microsoft Azure	20 CPU cores
Rackspace	128 GB memory
SoftLayer	20 compute instances per day, additional instances reviewed manually for approval

Default Quotas per Service

Instance Size	Amazon EC2	DigitalOcean	Google	Microsoft Azure	Rackspace	SoftLayer
1 core / 2 GB	160 (20 per region)	10	72 (24 per region)	20	64	20 daily
2 Cores / 4 GB	160 (20 per region)	10	32 (12 per region)	10	32	20 daily
4 Cores / 8 GB	160 (20 per region)	10	18 (6 per region)	5	16	20 daily
8 Cores / 16 GB	160 (20 per region)	10	9 (3 per region)	2	8	20 daily
16 Cores / 32 GB	160 (20 per region)	10	3 (1 per region)	1	4	20 daily

Storage Offerings

There are two types of disk volumes for compute services: local and external. Local storage volumes are located in the same physical hardware as your compute instance, while external volumes reside on an external system. The following table summarizes the pros and cons of local and external storage.

Local vs. External Storage Pros and Cons

Storage Type	Pros	Cons
Local	• Often faster and more consistent because it is not networked	• Less durable - hardware failure may result in loss of both compute instance and data • If you terminate the compute instance, storage volumes are lost • Often a fixed size depending on the compute instance type (the hardware can only have so many drives) • Usually doesn't support advanced features like volume backups/snapshots, copying or migration to another compute instance
External	• More scalable - size and quantity of volumes is flexible • Often supports advanced features like volume backups/snapshots, copying, volume sharing, or migration to another compute instance • Independent of the compute instance • More durable - failure of compute instance does not result in loss of a volume	• Often slower and less consistent due because it is network connected • Performance may be affected by changing network conditions • Although built to be more fault tolerant than a hard drive - failure may occur • Sometimes providers charge additional fees for input/ouput (IO) commands

Provisioned IO

Because external storage is networked, it can be prone to higher performance variability than local storage. Some workloads require consistent and predictable IO, which can be difficult to achieve with external storage. To address this problem, some services offer (or use exclusively) a provisioned IO model. With provisioned IO, you select the amount of IO capacity you need, and you are guaranteed to always get that capacity. With this approach, your IO performance will be highly consistent as opposed to standard networked external storage where you IO may be highly variable and unpredictable.

Service Storage Capabilities

The following table summarizes the storage capabilities for each compute service. We listed types of disk drives providers use if the provider discloses this information. Providers that include the Snapshots capability allow you create a backup of a storage volume, which you can use to create new volumes.

Compute Service	Local	External	Drive Types	Multiple Volumes	Snapshots	Provisioned IO
Amazon EC2	Yes	Yes	SATA SSD	Yes	Yes	Yes
DigitalOcean	Yes	No	SSD	No	Yes	No
Google	Beta	Yes	Undisclosed	Yes	Yes	Yes
Microsoft Azure	Yes	Yes	Undisclosed	Yes	Yes	No
Rackspace	Yes	Yes	SATA SSD	Yes	Yes	No
SoftLayer	Yes	Yes	SATA	Yes	Yes	No

Networking Capabilities

All compute services support basic network capabilities allowing you and your users connect to compute instances using an Internet Protocol version 4 (IPv4) address. Some common additional capabilities are support for IP version 6 (IPv6), and support for multiple IPv4 addresses, private networks, load balancing, and health checks.

IPv6 Support

Accessing a compute instance requires that the instance have a unique IP address. IPv4 is the current standard addressing protocol for the Internet. Because of a shortage of IPv4 addresses, however, IPv6 is being slowly adopted as a replacement.

Multiple IPv4 Addresses

Because of the IPv4 address shortage, many providers allow you to only have one address per compute instance. This is an issue if you need to support multiple Secure Socket Layer (SSL) certificates on a single compute instance, because each certificate requires a separate IP address. There are some workarounds,but these may or may not be feasible for you.

Private IP Address

It is also useful to have a separate network for connecting your compute instances in the same data center. For example, your web and database servers might communicate over a separate IP address without disrupting the Internet network. Communicating in this way is often faster and also more secure because the network traffic never crosses a publicly visible link. This may also reduce bandwidth costs, because private network bandwidth usually isn't metered.

Load Balancing and Health Checks

Load balancing allows you to spread incoming traffic across multiple compute instances. Load balancing is useful for scalability and durability because it distributes load across multiple compute instances, you can often configure automatic failover in case an instance fails.

Service Networking Capabilities

The following table summarizes the networking capabilities of each compute service.

Compute Service	IPv6 Support	Multiple IPv4	Private IP	Load Balancing	Health Checks
Amazon EC2	Yes	Yes	Yes	Yes	Yes
DigitalOcean	No	No	Partial (3 of 6 regions)	No	No
Google	No	No	Yes	Yes	Yes
Microsoft Azure	No	Yes	Yes	Yes	Yes
Rackspace	Yes	Yes*	Yes	Yes	Yes
SoftLayer	Yes	Yes	Yes	Yes	Yes

* Must request with support ticket with Rackspace and provide valid justification

Data Center Locations

Compute services often have data centers in multiple locations and allow you to decide which you want to deploy compute instances to. If you know where your most of your users are located, you can place your compute instances close to your users, so the instances can be more responsive over shorter network routes. You might also have users who are distributed globally, and want to use multiple servers in each region to support them. Number of data centers may also indicate the scale at which a provider operates.

The following table lists the number of data centers or regions that each compute service has in five continents.

Compute Service	North America	South America	Europe	Asia	Australia
Amazon EC2	4	1	1	3	1
DigitalOcean	3	0	1	1	0
Google	1	0	1	1	0
Microsoft Azure	6	1	2	4	0
Rackspace	4	0	1	1	1
SoftLayer	4	0	1	1	0

Security Features

Security features allow you to restrict access to resources and protect data. Some common security features include firewalls, Virtual Private Networking (VPN), Virtual Private Clouds (VPC), and Payment Card Industry (PCI) Data Security Standards (DSS) compliance.

Firewall

Firewalls are the most common and useful security feature. Services using firewalls keep the Internet IP address of your compute instances on a separate firewall device, and let you configure the access you want to allow. Most operating systems include a software firewall, but blocking traffic from an external device specialized for this purpose is usually a better and simpler approach. Firewalls can also make it easier to manage security by controlling access to multiple compute instances with a single set of access rules.

VPN

A VPN allows you to establish secure connectivity between your compute instances and an offsite client or network. For example, you might want to connect your office network to the compute service. A VPN automatically encrypts all traffic between the two locations.

VPC

A VPC enables your compute instances to communicate over physically or logically isolated networks, meaning that other service users cannot see or sniff your traffic. This differs from the private IP feature discussed previously because private IP addresses might still be visible by other users of the service.

PCI DSS Compliance

PCI DSS is a certification of compliance by the payment card industry, and is usually required to process credit cards electronically. Providers must adhere to certain guidelines related to facility access and security in order to claim compliance. Keep in mind that just because the service is compliant does not mean your software is: you must still adhere to additional standards in your software in order to be fully compliant. If the compute service is not compliant to begin with, however, you cannot be fully PCI DSS compliant, and might need to rely on an external service for credit card processing.

Service Security Features

The following table summarizes the security features of each compute service.

Compute Service	Firewall	VPN	VPC	PCI DSS
Amazon EC2	Yes	Yes	Yes	Yes
DigitalOcean	No	No	No	No
Google	Yes	No	Yes	No
Microsoft Azure	Yes	Yes	Yes	Yes
Rackspace	$160/mo per server	$160/mo per server	Yes	Yes
SoftLayer	Yes	Yes	Yes	Yes

* Requires additional subscription to Brocade Vyatta vRouter starting at $160 per month for each compute instance

Service Ecosystem

Cloud compute is one of many types of cloud services. Other cloud services include object storage, Content Delivery Networks (CDN), Domain Name System (DNS), database-as-a-service (DBaaS), and platform-as-a-service (PaaS). Often your workloads may need to use more than one of these services. If this is the case, you should consider the total service ecosystem of each provider because it is usually easier to work with only one vendor both from an administrative and implementation perspective. The sections below summarize each of these additional cloud service types and whether or not the providers included in this report support them.

Object Storage

Storing files on a compute instance disk allows only that compute instance to access them. Sometimes you might need to enable shared access to files. An object storage service is specifically designed for this purpose. Storing files on an object storage service allows multiple users or compute instances to access files at the same time. Object storage services often support higher durability by replicating files across multiple drives and data centers, and improve security by allowing you to define custom access policies. Object storage services are also accessible over standard web ports, making it much easier to grant access to them.

CDN

A CDN service allows you to host web content from a fully managed and globally distributed network of servers known as points-of-presence (POPs). A CDN can improve web performance significantly by responding to user requests using a POP that is in closer proximity to the user than your own web server. For example, if you had a web server in Europe, but your users were in Australia, your users would probably experience a dramatically better performance from a CDN POP in Australia versus your web server. CDNs are often used for faster delivery of static web content (some services even support dynamic content). CDN performance is dependent on the location and number of POPs a service uses, and the method used for POP selection.

DNS

A DNS service allows you manage IP addresses used for a hostname like www.example.com. A DNS service, like a CDN, uses a globally distributed network of servers, and performance also depends on the location and number of POPs the service uses.

DBaaS

A DBaaS is a fully managed database service for relational or NoSQL servers such as MySQL, PostgreSQL and MongoDB. Databases are hosted in multitenant (shared) or single-tenant (dedicated) environments. A DBaaS simplifies your use of a database by taking care of all of the administrative tasks. Databases are configured, patched, upgraded, and monitored by the providers' administrators. Single-tenant services are often abstraction layers built on top of a compute service.

PaaS

A PaaS allows you to deploy complete applications to the cloud without setting up, configuring, or managing servers. A PaaS often supports a full suite of capabilities for applications including web and database servers. A PaaS simplifies management of applications by removing the overhead associated with managing servers, scaling, failover, and many other administrative tasks.

Other Services Supported

The following table lists the other types of cloud services supported by each provider. The POP counts shown for CDN and DNS services are based on the number of unique locations in each continent listed, that is, a single location with multiple POPs is counted as one.

Compute Service	Object Storage	CDN	DNS	DbaaS	PaaS
Amazon EC2	Yes	Yes POPs: North America: 14 South America: 2 Europe: 10 Asia: 9 Australia: 1	Yes POPs: North America: 14 South America: 2 Europe: 10 Asia: 9 Australia: 1	Relational - Single-tenant Key/Value - Multitenant	Yes
DigitalOcean	No	No	No	No	No
Google	Yes	No	Yes POPs unknown	Relational - Multitenant Key/Value - Multitenant	Yes
Microsoft Azure	Yes	Yes POPs: North America: 8 South America: 1 Europe: 7 Asia: 7 Australia: 1	No	Relational - Multitenant	Yes
Rackspace	Yes	Yes - resells Akamai CDN. Uses 219 POPs globally	Yes POPs: North America: 2 Europe: 1	Relational - Multitenant	No
SoftLayer	Yes	Yes - resells Edgecast	Yes POPs: North America: 11 Europe: 3 Asia: 3	No	No

Conclusion

This report has tackled some questions we've found to be common among decision makers attempting to choose between cloud providers:

- Which cloud provider should I choose?
- How do I compare performance across cloud services?
- Is there a fastest cloud?
- Which provider is best for my workloads?
- Which provider offers the best value?

We've shown that because cloud providers use dissimilar terms to describe performance and rely on oversubscription, sometimes excessively, measuring and comparing cloud performance comes with some unique challenges.

Reliable comparisons of cloud services are hard to find. Existing studies are often flawed by irrelevant benchmarks, incorrect text execution, or lack of understanding of the nuances of cloud. These flawed studies contribute to flawed claims and headlines about how services compare with one another.

We hope this report has provided you with some useful knowledge and techniques for testing and comparing performance of cloud services, and that you are better prepared to choose a provider with confidence using this knowledge.