



Real experts.  
Real data.  
Real savings.

WHITE PAPER

# The Definitive IT Sourcing Guide to Cloud Vendor Competitiveness, Pricing and Support

**PART III: HOW DO VENDOR SERVICE LEVEL  
AGREEMENTS AND SUPPORT STACK UP?**

*Featuring comparisons among Amazon Web Services,  
Microsoft, Google, IBM and Oracle*

## Table of Contents

SUPPORT .....	3
AWS .....	3
MICROSOFT .....	4
ORACLE.....	4
IBM .....	5
GOOGLE .....	5
SERVICE LEVEL AGREEMENTS .....	7
AWS .....	8
MICROSOFT .....	8
ORACLE.....	9
IBM .....	9
GOOGLE .....	10
SLA COMPLIANCE .....	11
OVERALL OBSERVATIONS .....	13
THE SWEET SIXTEEN: TIPS FOR PROTECTING YOUR CLOUD INVESTMENT .....	13

## About This Series

The cloud may be pervasive, but doing business with cloud vendors (particularly IaaS/PaaS) is still immature in many ways. What are each vendor's strengths and weaknesses? What pricing nuances should be considered? And how does support compare?

From an IT Sourcing perspective, this three-part white paper series explores what really sets five IaaS/PaaS vendors apart in the following three areas:

Part I: Competitiveness

Part II: Pricing

**Part III: Support**

In the first two installments of this white paper, we compared cloud provider *performance* and *pricing*. In this final part, we compare the service level agreements (SLAs) and support.

Many enterprises are surprised at how cloud service providers handle accountability – especially those accustomed to traditional IT outsourcing contracts. AWS, Microsoft and Google are relatively new to the IT service provider space, and their services are designed to appeal horizontally across as broad a market as possible. Buying IaaS is more akin to buying software than buying fully managed services even though service commitments and support tend to be more generic than what you find with managed services.

## Support

For support, you have a variety of standard menu options from each provider, and each higher tier includes the features of all the lower ones.

AWS	LEVEL 1	LEVEL 2	LEVELS 3 & 4	LEVEL 5
<b>SUPPORT TYPE</b>	Basic	Developer	Business	Enterprise
<b>DESCRIPTION</b>	<ul style="list-style-type: none"> <li>• Support forums</li> <li>• 6 “core” Trusted Advisor checks</li> <li>• Health dashboard</li> <li>• Documentation</li> </ul>	<ul style="list-style-type: none"> <li>• Web/email support for one user</li> <li>• Best-practice guidance</li> <li>• Client-side diagnostic tools</li> <li>• Building-block architecture support: guidance on how to use AWS products, features, and services</li> </ul>	<ul style="list-style-type: none"> <li>• 24x7 phone/chat support for multiple users via IAM</li> <li>• Service selection guidance</li> <li>• Full Trusted Advisor</li> <li>• Support Center API</li> <li>• EC2 OS instance configuration</li> <li>• Performance support</li> </ul>	<ul style="list-style-type: none"> <li>• Sr. Support Engineers</li> <li>• Guidance on architecture and scaling</li> <li>• Short-term event management</li> <li>• Technical Account Manager</li> <li>• Priority escalation</li> <li>• Management/operations reviews</li> </ul>
<b>INCIDENT RESPONSE TIME</b>	N/A	<ul style="list-style-type: none"> <li>• System impaired: &lt;12 bus. hours</li> <li>• General guidance: &lt;24 bus. hours</li> </ul>	<ul style="list-style-type: none"> <li>• Production down: &lt; 1 hour</li> <li>• Production impaired: &lt; 4 hours</li> <li>• System impaired: &lt; 12 hours</li> <li>• General guidance: &lt; 24 hours</li> </ul>	<ul style="list-style-type: none"> <li>• Business-critical system down: &lt; 15 minutes</li> </ul>
<b>PRICING</b>	Free	<ul style="list-style-type: none"> <li>• \$29 minimum</li> <li>• 3% of bill</li> </ul>	<ul style="list-style-type: none"> <li>• \$100 minimum</li> <li>• 10% of bill for the first \$0–\$10K</li> <li>• 7% of bill from \$10K–\$80K</li> <li>• 5% of bill from \$80K–\$250K</li> <li>• 3% of bill &gt; \$250K</li> </ul>	<ul style="list-style-type: none"> <li>• \$15K minimum</li> <li>• 10% of bill for the first \$0–\$150K</li> <li>• 7% of bill from \$150K–\$500K</li> <li>• 5% of bill from \$500K–\$1M</li> <li>• 3% of bill &gt; \$1M</li> </ul>

MICROSOFT	LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4	LEVEL 5
<b>SUPPORT TYPE</b>	Included	Developer	Standard	Pro-Direct	Premier <sup>†</sup> + Rapid Response
<b>DESCRIPTION</b>	<ul style="list-style-type: none"> <li>Forums and Twitter</li> <li>Resource Health tool</li> <li>Documentation</li> </ul>	<ul style="list-style-type: none"> <li>Online support only</li> </ul>	<ul style="list-style-type: none"> <li>Includes Azure Stack</li> <li>24x7 phone support for Severity A incidents</li> <li>Optional for Severity B</li> </ul>	<ul style="list-style-type: none"> <li>Proactive monitoring and escalation for of high-severity incidents</li> <li>Account management team</li> <li>Best practice guidance</li> <li>Monthly service reviews with limited advisory</li> </ul>	<ul style="list-style-type: none"> <li>Includes all MS products, not just Azure</li> <li>Assigned Account Manager</li> <li>Onsite support</li> <li>Best practice support on architecture, code, deployment, performance, security, hybrid, reliability, recoverability</li> <li>Preventative checks</li> </ul>
<b>INCIDENT RESPONSE TIME</b>	N/A	<ul style="list-style-type: none"> <li>All incidents treated as Severity C "minor impediments:" &lt; 8 bus. hours</li> </ul>	<ul style="list-style-type: none"> <li>Severity A: &lt; 2 hours</li> <li>Severity B: &lt; 4 bus. hours*</li> <li>Severity C: &lt; 8 bus. hours</li> </ul>	<ul style="list-style-type: none"> <li>Severity A: &lt; 1 hours</li> <li>Severity B: &lt; 2 bus. hours*</li> <li>Severity C: &lt; 4 bus. hours</li> </ul>	<ul style="list-style-type: none"> <li>Severity A: &lt; 15 minutes</li> <li>Severity B: &lt; 2 bus. hours*</li> <li>Severity C: &lt; 4 bus. hours</li> </ul>
<b>PRICING</b>	Free	• \$29/month	• \$300/month	• \$1,000/month	Negotiated

ORACLE	LEVELS 1 & 2	LEVELS 3-5
<b>SUPPORT TYPE</b>	Oracle Cloud Support	Advanced Customer Support
<b>DESCRIPTION</b>	<ul style="list-style-type: none"> <li>24x7 web portal and phone support for designated contacts</li> <li>Community forums</li> </ul>	<ul style="list-style-type: none"> <li>Includes all Oracle products, technical account manager, customer-specific support portal, priority escalation, migration, workload planning, consolidation, launch readiness, quarterly progress reports, etc.</li> </ul>
<b>INCIDENT RESPONSE TIME</b>	N/A	<ul style="list-style-type: none"> <li>Severity 1: 90% in 15 minutes</li> <li>Severity 2: 90% in 2.5 business hours</li> <li>Severity 3/4: 90% by next business day</li> </ul>
<b>PRICING</b>	Free	Negotiated Pricing

\*24x7 available for extra cost. †At time of publishing, it's been reported that Microsoft may replace Premier Support with Unified Support. To learn more about how these changes could impact Azure support, please contact NPI.

IBM	LEVELS 1 & 2	LEVEL 3	LEVELS 4 & 5
<b>SUPPORT TYPE</b>	Free	Standard	Premium
<b>DESCRIPTION</b>	<ul style="list-style-type: none"> <li>• Support forums</li> <li>• Status page</li> <li>• Support portal</li> <li>• Documentation</li> </ul>	<ul style="list-style-type: none"> <li>• 24x7 phone support for Severity 1 incidents</li> </ul>	<ul style="list-style-type: none"> <li>• Phone/chat support</li> <li>• Client Success Manager</li> </ul>
<b>INCIDENT RESPONSE TIME</b>	All incidents treated as Severity 4	<ul style="list-style-type: none"> <li>• Severity 1: &lt; 1 hour</li> <li>• Severity 2: &lt; 2 bus. hours</li> <li>• Severity 3: &lt; 4 bus. hours</li> <li>• Severity 4: &lt; 8 bus. hours</li> </ul>	<ul style="list-style-type: none"> <li>• Severity 1: &lt; 1 hour</li> <li>• Severity 2: &lt; 1.5 bus. hours</li> <li>• Severity 3: &lt; 2 bus. hours</li> <li>• Severity 4: &lt; 4 bus. hours</li> </ul>
<b>PRICING</b>	Free	• Negotiated – “starts at \$200/mo”	• Negotiated – “starts at \$10,000/mo”

GOOGLE	LEVEL 1	LEVELS 2 & 3	LEVEL 4	LEVEL 5
<b>SUPPORT TYPE</b>	Bronze	Silver	Gold	Platinum
<b>DESCRIPTION</b>	<ul style="list-style-type: none"> <li>• Support forums</li> <li>• Status dashboard</li> <li>• Documentation</li> </ul>	<ul style="list-style-type: none"> <li>• Web support for 2 users</li> <li>• Architecture best practice guidance</li> </ul>	<ul style="list-style-type: none"> <li>• 24x7 phone support for P1 incidents</li> <li>• Web support for 5 users</li> <li>• Architecture support</li> <li>• Consultation</li> </ul>	<ul style="list-style-type: none"> <li>• 24x7 phone support for P1/P2 incidents</li> <li>• Web support for unlimited users</li> <li>• Technical account management team**</li> </ul>
<b>INCIDENT RESPONSE TIME</b>	N/A	<ul style="list-style-type: none"> <li>• P1: &lt; 4 bus. hours</li> <li>• P2/3/4: &lt; 8 bus. hours</li> </ul>	<ul style="list-style-type: none"> <li>• P1: &lt; 1 hour</li> <li>• P2: &lt; 4 bus. hours</li> <li>• P3/4: &lt; 8 bus. hours</li> </ul>	<ul style="list-style-type: none"> <li>• P1: &lt; 15 minutes</li> <li>• P2: 4 hours</li> <li>• P3/4: &lt; 8 bus. hours</li> </ul>
<b>PRICING</b>	Free	• \$150/month	<ul style="list-style-type: none"> <li>• \$400 minimum</li> <li>• 9% of bill for the first \$0–\$10K</li> <li>• 7% of bill from \$10K–\$50K</li> <li>• 5% of bill from \$50K–\$200K</li> <li>• 3% of bill &gt; \$200K</li> </ul>	<ul style="list-style-type: none"> <li>• \$15K minimum</li> <li>• 9% of bill for the first \$0–\$150K</li> <li>• 7% of bill from \$150K–\$400K</li> <li>• 5% of bill from \$400K–\$1M</li> <li>• 3% of bill &gt; \$1M</li> </ul>

\*\*Extra charges may apply.

Here are a few things to consider as you compare support options: Response time targets listed here are not backed up by financial incentives, so they should just be understood as the goals the service provider uses to design the service. Support in languages other than English may require reduced hours, a more expensive support plan or may not be available at all. As for severity levels, these are initially set by the customer when a ticket is opened, but providers can reduce those levels at their discretion.

AWS Developer, Business and Enterprise support require a monthly minimum, billed in advance. Microsoft Standard, Pro-Direct and Premier support can be bundled into an EA. While phone numbers are available, Microsoft actively discourages support phone calls and recommends their support portal for all levels of support. If you'd prefer a fully managed service, each of these providers has partners that will manage the service for you and leverage their own support relationship with the cloud service provider.

Overall, there are a few conclusions you can draw from the tables above:

- Support pricing for AWS, Azure and GCP are similar but not exactly the same. Microsoft has 5 optional plans instead of 4, and their Professional Direct support appears to be aggressively priced at a flat \$1000 per month. If that support level meets your needs you can benefit from substantial savings vs. AWS Business or Google Gold support.
- Large enterprises and organizations putting mission critical systems in the cloud are likely to prefer the highest support option available from any of these providers. The only provider without a 15-minute response target for critical incidents is IBM.
- Microsoft and Oracle can both address incidents without finger pointing when it's unclear if the cloud platform or their own software running on the platform is the culprit, and that can be a significant benefit, depending on which products you use.
- Microsoft, IBM and Oracle have negotiated pricing for their highest level of support, while AWS and Google stick to a standardized pricing menu. If you have large support contracts for other products with the first three, you should be able to benefit by leveraging those when you negotiate your cloud support. Since you're far less likely to have that for AWS or Google, it makes sense that they stick with generic pricing. Note that 3% of your bill is quite a bit less than what software companies have traditionally charged for maintenance and support, which is typically 18 percent or more.
- Oracle is the maverick here. You either get the provider's free support, which pushes forums and portal-based channels but does include phone support, or you go with an enhanced service that is negotiated and priced similarly to Oracle's software products with numerous options. On the low end, it has the cheapest phone support of any provider. On the high end, Oracle has very extensive consultative support that at least rivals the competition, though pricing is not published and cost could be significant.

In a survey of more than 624 IT and Sourcing professionals, respondents were asked: Do you plan to expand your use of IaaS/PaaS in 2018?



SOURCE: NPI CLIENT SURVEY, APRIL 2018

- 46.66% Yes, a lot
- 40% Yes, a little
- 6.67% No, keeping the status quo
- 6.67% No, plan to shrink usage

## Service Level Agreements

SLAs for public cloud services are not similar to SLAs for traditional IT outsourcing services. Credits tend to be generous, but the potential for actually receiving one is low, even though they've gotten a little stronger recently. Here's a breakdown of what you can expect.

PROVIDER	SERVICE	MONTHLY TARGET	FEE CREDIT
AWS	EC2 (including EBS)	99.0% to < 99.99%	10%
		< 99.0%	30%
	S3	99.0% to < 99.9%	10%
		< 99.0%	25%
	S3 Standard Infrequent Access	98.0% to < 99.0%	10%
		< 98.0%	25%
	RDS Multi-AZ instances other than SQL Server	99.0% to < 99.95%	10%
		< 99.0%	25%
	Most other AWS services, including Lambda	None	None

PROVIDER	SERVICE	MONTHLY TARGET	FEE CREDIT
MICROSOFT	Azure VMs in availability set	99% to < 99.95%	10%*
		95% to < 99%	25%*
		< 95%	100%*
	Azure single instance VMs using only premium storage	99% to < 99.99%	10%*
		95% to < 99%	25%*
		< 95%	100%*
	Storage – LRS, ZRS, GRS and RA-GRS (write requests)	99% to < 99.9%	10%*
		< 99%	25%*
	Read requests from Read Access, Geographically Redundant Storage	99% to < 99.99%	10%*
		< 99%	25%*
	Storage – Infrequent Access (“Cool Access Tier”)	98% to < 99%	10%*
		< 98%	25%*
	Storage – Infrequent Access, Read requests from RA-GRS	98% to < 99.9%	10%*
		< 98%	25%*
	SQL Database – Basic, Standard, Premium	99% to < 99.99%	10%*
		< 99%	25%*
Azure Functions - Consumption Plan	None	None	
Numerous Other Azure Services	Various	Various	

\*Percentage applies only to fees for the affected resource.



PROVIDER	SERVICE	MONTHLY TARGET	FEE CREDIT
ORACLE	Category 1, 3, 4, 5 Services, including compute, block volumes, database	< 99.95%	None
	Category 2 services, including Backup, Object Storage and Infrastructure Network	< 99.9%	None

PROVIDER	SERVICE	MONTHLY TARGET	FEE CREDIT
IBM	Platform Services - High Availability, Multiple Region OR Multiple Dedicated Local Locations	99.0% to < 99.95%	10%
		< 99.0%	25%
	Multiple Instances in Single Dedicated or Local Environment	99.0% to < 99.5%	10%
		< 99.0%	25%
	Public or Private Network, HVAC	5% for each 30 minutes per outage	5% for each 30 minutes per outage
	Hardware replace/upgrade response	> 2 hours to 6 hours	20%
		> 6 hours to 10 hours	40%
		> 10 hours to 14 hours	60%
		> 14 hours to 18 hours	80%
		> 18 hours	100%
	Most other Cloud Services, including Cloud Functions	None	None

PROVIDER	SERVICE	MONTHLY TARGET	FEE CREDIT
GOOGLE	Google Compute Engine (VMs and load balancing)	99.00% to < 99.95%	10%
		95.00% to < 99.00%	25%
		< 95.00%	50%
	Multi-regional Storage	99.0% to < 99.95%	10%
		95.0% to < 99.0%	25%
		< 95.0%	50%
	Regional Storage	99.0% to < 99.9%	10%
		95.0% to < 99.0%	25%
		< 95.0%	50%
	Nearline, Coldline, Durable Reduced Availability Storage	98.0% to < 99.0%	10%
		95.0% to < 98.0%	25%
		< 95.0%	50%
	Cloud SQL (Google Cloud SQL 1st gen., MySQL, PostgreSQL)	99.0% to < 99.95%	10%
		95.0% to < 99.0%	25%
		< 95.0%	50%
Cloud Functions (Beta)	None	None	
Numerous Other GCP Services	Various	Various	

Note that SLA compliance management is not proactive for any of these providers. You need to determine that there is a problem and report it, and then the provider must agree, before you will be issued a credit. The time you have to submit your claim is limited, usually to no more than two months, and if one event causes SLA non-compliance for multiple services, you must file a separate claim for each affected service.

Keep in mind that uptime is calculated across all VMs, just as it is with traditional ITO, so having only a few VMs down for a long time will not result in a credit if there are enough other VMs running to raise the average above the uptime target. It is not unusual for services to be provided without any SLA at all (e.g., Function as a Service), and the SLAs that are provided typically end at an uptime commitment; degraded service won't result in any fee credits. VMs with any external connectivity are considered "up."

### AWS

- The only downtime that qualifies is "regional," which means that more than one availability zone used by you in the same region is down. If you don't use more than one availability zone, you have no SLA.
- Outages that "result from failures of individual instances or volumes not attributable to Region Unavailability" are specifically excluded.
- S3 uptime is determined by sampling the error rate for 5-minute intervals and taking the average over the entire month. That average is subtracted from 100%. That means your service can be down for a long period of time (0.1% of a month is about 44 minutes) and not trigger a fee credit as long as the error rate is minimal for the rest of the month.
- Fee credits are against the bill for the entire service (EC2, EBS, etc.), not just the bill for the affected resources. One-time charges, including any up-front payments made for Reserved Instances, are excluded.

### MICROSOFT

- Although the uptime calculation for Azure VMs is an average across the entire service (all your VMs), the fee credits apply only to charges for the "service resource" that had the outage (individual VMs). Considering that those individual VMs must have had significant downtime to trigger the SLA non-compliance, if you're on a pay-as-you-go plan the charges the credit is calculated against could be pretty small, and that's going to minimize your credit.
- Outages less than one minute long do not count as downtime for VMs.
- Storage uptime is calculated using error rates, as with AWS S3, but an "error" is any request that exceeds a specified maximum time, and those can be quite long, so you can experience significantly degraded performance and still be in compliance with the SLA. For Read Access Globally Redundant Storage, your application must retry any failed request using the redundant copy before it will count as an "error."
- Microsoft does offer an SLA for Azure Functions if you opt for the "App Service" Plan instead of the "Consumption" Plan. An App Service Plan gives you a dedicated VM instance to run your functions on instead of providing the automatic scaling that you get with the Consumption plan or AWS Lambda. Unsurprisingly, that SLA is similar to the ones they offer for VMs.

When asked if they could change anything about their experience with IaaS/PaaS providers, respondents expressed concern around pricing and support transparency as well as providers' ability to be a strategic partner in solving critical business problems. Sample responses include:

*"Pricing, add-on costs, technical support, and better alignment to organization's needs instead of the vendor's."*

*"A more strategic approach to helping us solve business problems."*

*"Cost is still more expensive than on-premise."*

*"Ability to understand the full cost picture, including the IaaS/PaaS services and wrap-around services."*

*Source: NPI Client Survey, April 2018*

## ORACLE

- Oracle Cloud Services are grouped into 6 different categories for the purpose of Service Level Objectives, though several categories carry the same uptime target. Commitments are at the level of the overall service and not individual resources, and connectivity must be unavailable for the entire service to qualify as “downtime.”
- Category 2 objectives measure uptime for object storage by subtracting an average error rate for the month from 100%, similar to the Storage SLAs from the other providers. Uptime for block storage is in Category 5, however, and is simply measured by the percentage of time you can connect to it, the same way that uptime is measured for VMs.
- Note that Oracle uses the term “objectives” when describing service level targets. There are no fee credits associated with these targets, possibly reflecting the relative youth of the offerings.

## IBM

- Downtime means that no instances can be connected to for the platform service. If even one is up, it’s not downtime.
- SLA credits must be “claimed” within only 7 days of the end of a billing period after an outage. This is the customer’s responsibility.
- IBM structures SLAs somewhat differently, adding targets and credits for hardware replacements and upgrades, and since it specializes in bare metal, dedicated hardware that each client can upgrade. IBM also covers dedicated equipment provided locally to the client.
- Most cloud services beyond the actual compute instances and the network do not have SLAs.

## GOOGLE

- Similar to AWS, the downtime referred to is only for “all running instances hosted across two or more zones in the same region.” If one instance in those zones is still up, or if you don’t use more than one availability zone, you have no downtime as far as the SLA is concerned.
- Outages less than one minute long do not count as downtime for VMs.
- Google Cloud Storage downtime is calculated with error rates, but duplicate requests for storage access do not count as errors unless they conform to “back-off” requirements, which require the application to wait for periods beginning at 1 second and ramping up to 32 seconds before retrying the request. This acts to reduce the number of errors being used to calculate the error rate.
- Google Cloud SQL downtime is also measured with an error rate, similar to storage, that must be at least 20% based on at least 5 attempts in 1 full minute. To count as downtime for MySQL and PostgreSQL, all requests must fail for 1 minute to a multi-zone instance, which means that the dedicated core CPU instance must have a failover replica.
- At the time this paper was published, Google Cloud Functions is not covered by an SLA because it is still a beta release. Once a full production release is delivered an SLA may or may not be provided.

## OVERALL OBSERVATIONS

Overall, Microsoft and Google probably have the strongest SLAs, just based on the extensive list of services that are covered. AWS and IBM provide them only for a limited list of core service offerings, which means that downtime caused by any new, app-specific or ancillary services of the type that are often needed for real-world applications isn't going to be covered. Oracle is tracking all the metrics required for an extensive set of SLAs, but they simply haven't announced any solid financial commitments yet.

Overall, SLAs in the public cloud market are consistently abysmal. Where they exist, they are crafted to make sure that your chances of ever claiming a fee credit, even for substantial outages, are very, very low, even when customers are vigilant. The providers know that customer cloud governance tends to be weak, and placing the burden of SLA compliance monitoring on them lowers the odds of financial impact. Customers that value SLAs must be prepared to put robust governance in place if their SLAs are to have any impact at all, but expectations should still be relatively low.

Uptime KPIs are the only ones receiving significant attention in the public IaaS market thus far. There are also commitments to response times for support but without any financial incentives to back them up. If you want an SLA for transaction response time, elasticity, problem resolution time, customer satisfaction or any of the other KPIs you may have been accustomed to using with traditional ITO, you're probably going to be disappointed. Until customers begin demanding these additional KPIs, we probably won't see much movement from the providers.

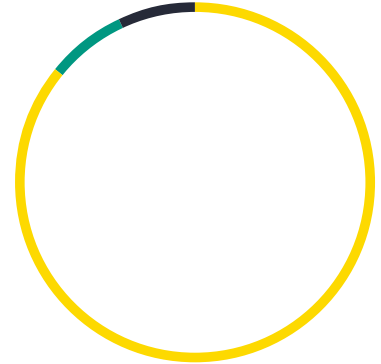
Customers that value SLAs must be prepared to put robust governance in place if their SLAs are to have any impact at all, but expectations should still be relatively low.

## The Sweet Sixteen: Tips for Protecting Your Cloud Investment

- **Put cloud-specific governance in place.** The list of customers that have put their first application or two in the cloud and then not realized the savings they expected is very long, and lack of proper governance is a big reason for that. Providers will give you basic tools for managing things like unused resources, but it's up to you to use those tools effectively. Remember, in a pay-for-use world, anything you don't turn off when you don't need it is costing you money. Add to this the fact that application workloads are virtually never static. The financially optimal configuration you're using today, even the optimal service you're using today, may not still be optimal tomorrow. Developers are notorious for leaving resources allocated when they don't need them anymore (commonly referred to as "zombies"), and processes need to be put in place for continuous zombie search and destroy.
- **Remember, not all data requires the same performance level.** Stored data likes to grow and grow, but old data may not need the same performance levels as new data, and the cloud offers many different price points for storage. That means HSM and ILM techniques are more important than ever. When public cloud providers lower prices, they often do so on new configurations only. If your apps are still running on old generation instance types you could be throwing money out the window every minute that you leave them there. You can avoid some of this with money-saving options like AWS's Convertible Reserved Instances and Spot Market Instances, but even those require management effort to realize their benefits. When your cloud presence is big enough to justify it, look at Cloud Management Platforms that can help you manage all this effectively.
- **Use cloud-aware applications to release your resources automatically.** The governance tasks we just mentioned are that much harder if your applications aren't helping. You want your applications to release your compute instances and purge or archive your old data to cheaper storage when you don't need it anymore. The same technology that the cloud provides for rapidly scaling up can also help you rapidly scale down to save money. This is the benefit of rapid "elasticity," and one of the core benefits of cloud computing, but your application has to be smart enough to use it.

- **Negotiate when you have leverage and when your “asks” won’t break the providers’ business model.** All of these providers, but especially AWS, Microsoft and Google, focus on selling through standardized catalogs with standardized discount structures, and it’s very difficult to move them away from those. While NPI is seeing some providers behave more flexibly at the negotiation table, you’ll have more luck negotiating things like SLA fee credits, free training and reduced charges for support or proof-of-concept projects (this is true for SaaS providers as well). As always, bigger financial commitments and more potential for publicity will motivate the providers to be more flexible. AWS went to significant lengths to win the CIA’s business, for example, including adding custom security features and running their “public” cloud infrastructure in the customer’s own data centers.
- **If you’re a Microsoft or Oracle customer, get help with complex licensing options if you need it.** Both of those firms may offer you multiple paths to buying cloud services, and at least some will be entwined with other software licensing you already have in place. Choosing the right path can save a lot of money, and analyzing the choices can be quite complex. Engage someone with experience to do the analysis, and don’t forget to project your costs over multiple years instead of just looking at what your near-term costs will be. Discounts can often be temporary.
- **Anticipate cost “tipping points.”** Customers that have gone fairly far down the cloud path report step-wise savings that can occur when an entire function, process or asset is made irrelevant by cloud. This could be a data center that is no longer needed, or something like a license for backup software (along with the expertise in operating it) that you no longer need because the data is now being backed up automatically in the cloud.
- **Do a real evaluation.** Success in sourcing cloud services is not as simple as just choosing the provider with the most market share or the one that your peer at another firm is happy with. And with public cloud you no longer have the luxury of specifying exactly what you need and expecting the providers to customize the service for you. As we’ve seen in this white paper series, there are dramatic differences from one provider to the next, both in the services they can offer and the prices they can offer them for. Choosing the wrong service can have a significant impact on the return you’ll realize on your project’s investment.
- **Test, or at least model, your applications.** Almost every aspect of your service, including what you’ll pay for it, depends on the characteristics of the application you are putting in the cloud. The workload you’re placing on the infrastructure, along with your need for support, will determine what the service will cost you, so you’ll want to test that in advance or at least model it as best you can. If you don’t do that, you have no idea what you will actually be paying.
- **In addition to engaging application developers in service selection, don’t forget to take factors like support, security, compliance, disaster recovery, interoperability, scalability, response time and ease of administration into account.** Not every provider has the same capabilities in these areas, and determining if the services meet your needs can require input and participation from a variety of experts on your team. The most expensive cloud project is the one that gets to completion and then fails because one of these factors wasn’t taken into account.
- **Leverage what you have.** Competition is still pretty fierce in the cloud market, as successes and failures in that space have an immediate impact on the provider’s stock price. If a Microsoft, IBM or Oracle is faced with losing a large cloud opportunity to an AWS or Google, they will definitely be in the mood to cut a deal. And if you already have a strategic relationship with one of them, you might avoid some retraining or re-staffing costs by taking their offer.

How would you rate the support you get from your IaaS/PaaS providers?



SOURCE: NPI CLIENT SURVEY, APRIL 2018

- 7.14% Excellent
- 85.71% Mostly acceptable
- 7.14% Mostly unacceptable

- **Don't just assume that pricing will always go down and SLAs will get better - pressure your providers.** We've seen price increases on specific configurations that get little to no coverage in the press but can make noticeable differences in your bill. And there's evidence that providers are not necessarily passing the cost benefits of Moore's Law on to customers, even while their profits are rising. That's because customers aren't demanding it strongly enough yet.
- **The highest returns long-term are usually from cloud-native applications, so prioritize those over "lift and shift."** There are many aspects of existing enterprise applications that can make them poor candidates for moving to the cloud, including short remaining life, fat-client design, dependency on specific OS, hypervisor, database or middleware versions and the inability to release resources automatically. Compelling business cases for lift and shift projects do exist, but developing cloud-native applications will usually deliver greater benefits assuming the development time and expense are practical. You may also find that having elastic cloud infrastructure makes entirely new application ideas feasible that may not have been so in the past, thus opening the door to new sources of revenue and productivity.
- **Consider a multi-cloud strategy.** It's becoming clear that the public cloud market, while it may have a well-known leader, is not going to simply be owned by a single provider. There's too much innovation going on in too many different areas for one provider to be best at everything. That's good because competition is required to hold prices down long-term and keep the market healthy. Large enterprises that can afford to do business with two strategic cloud service providers and still have a significant footprint with both will benefit from avoiding lock-in and delivering best of breed solutions for diverse applications. In addition to the cost benefits from choosing more optimal services for each application, there can be significant resiliency benefits, as one CSP can be used to recover from outages at the other.
- **If you're small, put an even higher priority on your move to cloud.** The economies of scale that the leading providers achieve are vast. Only large enterprises can compete with the price point of reserved instances from these providers, and even then it's a close contest.
- **Put a very high priority on testing, and assume there will be outages when you're designing the application.** Testing is even more important when you don't own the infrastructure and have to access it over the internet. In some cases, you're also dealing with fairly cutting-edge technology that might not always work as expected. All of the major providers have had significant outages, and SLAs are too weak to be considered a remedy or even an incentive to improve, so your application may need to be smart enough to failover automatically when it needs to. All of these factors, plus the ease of provisioning that cloud delivers, make cloud a great fit with a DevOps development methodology where development happens in relatively small chunks that are constantly being tested. If there is an unexpected problem, it's always better to find out earlier in your project rather than later.
- **Understand what you are outsourcing... and what you are not.** Public IaaS providers generally take responsibility only up to the OS, which they will license to you, but you still have to patch it and administer it just as you would if it were in your own data center. Security responsibilities that fall to the provider are relatively minor – they'll give you a firewall, wipe the storage that you stop using, keep the physical data center secure and background check their employees, but the ultimate responsibility for standards compliance, access management, virus scanning, OS hardening and many other security functions still resides ultimately with you. These are not fully managed services, and not all of your IT infrastructure labor cost is being eliminated when you use public cloud.



Real experts.  
Real data.  
Real savings.

**ABOUT NPI** — NPI is an IT sourcing consulting company that helps enterprises identify and eliminate overspending on IT purchases, accelerate purchasing cycles and align internal buying teams. We deliver transaction-level price benchmark analysis, license and service optimization advice, and vendor-specific negotiation intel that enables IT buying teams to drive measurable savings. NPI analyzes billions of dollars in spend each year for clients spanning all industries that invest heavily in IT. For more information, visit [www.npifinancial.com](http://www.npifinancial.com).