CS6204: Cumulus, Nimbus, Cumulonimbus – Choosing Between Clouds

Uday Ananth*, Jamie Davis*, and Ayaan Kazerouni*
*Virginia Tech, USA
Email: {uday, davisjam, ayaan}@cs.vt.edu

***Abstract*—Cloud computing is a multi-billion business, with more than 100 cloud service providers [1] vying for supremacy. Businesses are overwhelmed with choices and tradeoffs.**

In this paper we offer a qualitative and quantitative assessment of leading cloud service providers (CSPs) to guide small and medium businesses (SMBs) in their choice of CSP. Our qualitative study of usability, reliability, and customer service identified Azure as an intuitive and business-friendly cloud. In our quantitative work, we give SMBs some guidelines for drawing their own conclusions based on the performance characteristics of their applications.

***Index Terms*—Cloud computing, Performance, Benchmarking, Comparison, Usability study**

I. Introduction

Businesses rely on computing in many forms, from IT for daily business activities to HPC for number crunching and data analytics. In the past, businesses had to build, operate, and maintain their own computing infrastructure, at significant expense. The investment and skills required to obtain computational resources were once a significant barrier to entry and growth for SMBs [2].

Cloud computing has changed all of this. CSPs offer to manage computing resources for customers through offerings like Infrastructure-, Platform-, and Software-as-a-Service (IaaS, PaaS, and SaaS). Through the "pay-as-you-go" business model, a customer can obtain precisely the infrastructure, platform, or software services it needs, and pay for usage on a yearly, monthly, or even hourly basis. CSPs eliminate the up-front capital expense of purchasing equipment for customers and turn their operating expenses into an invoice. This offers SMBs cheaper and easier access to computing resources, allowing them to focus on their core business.

Though the "Big Four" of Amazon, Microsoft, IBM, and Google account for 51% of the market share [3], SMBs can pick from over 100 CSPs [1]. The decision is important, as the lack of standardization between CSPs means that SMBs risk becoming locked in to the CSP they choose [4]. We believe an SMB would consider both qualitative and quantitative characteristics of a CSP before making a decision. In this study, we evaluate three of the "Big Four" CSPs using both qualitative and quantitative methods: Amazon (AWS), Microsoft (Azure), and Google (Google).

II. Background and related work

Two primary frameworks exist to aid in evaluation of CSPs [5], [6]. We chose to study usability, reliability, customer service, and performance for each CSP; these characteristics represent a fuzzy intersection of the items included in these frameworks.

A. Qualitative

1) Usability: The usability of a system is a measure of how readily a user can accomplish his tasks. The usability of an interface can be evaluated in a number of ways. Using usability heuristics [7] or conducting empirical user-testing sessions ([8], [9]) are a few examples. While Zheng et al. [6] includes a cursory evaluation of the usability of CSPs' web interfaces, we were unable to find examples of thorough CSP usability evaluation.

2) Reliability: The reliability of a system describes the degree of variation in its availability and trustworthiness. Though work discussing the detection of SLA violations abounds (e.g. [10]–[12]), we could find no compilation of CSP outages in the scientific literature, nor any discussion of appropriate techniques to do so. CloudHarmony [13] measured outage rates by creating instances and pinging them for all of 2015, observing the number and length of connectivity outages that occurred. One drawback of this approach is that only outages that affect the instances under observation are detected.

3) Customer service: Customer service is a measure of how well customers are treated. Buyya et al. [2] emphasize the importance of customer service in cloud computing. However, they and others (e.g. [14], [15]) point to SLAs for uptime, security, and so on as the means to achieving this end. The CSPs in our study offer comparable SLAs [16]–[18], so customers cannot differentiate between them on that basis. We observe, however, that an SLA is meaningful only so long as a CSP meets it; when it does not the customer must work towards a solution with the CSP's support personnel. The findings of our reliability study suggest the need to consider the effectiveness of a CSP's support staff when choosing a CSP.

In their seminal work, Parasuraman et al. [19] propose the SERVQUAL framework to measure customer service quality (CSQ) along five dimensions: tangibles, reliability, responsiveness, assurance, and empathy. Applications of their approach abound, measuring CSQ in a variety of industries in a variety of ways. Early studies (e.g. [20], [21]) focused on CSQ in face-to-face settings. More recent work applied SERVQUAL in the context of internet-based services, including websites [22], [23], live chat [24], and email [25].

B. Quantitative

CSPs offer an alternative to traditional in-house computing resources, and their scale, complexity, and heterogeneity result in performance variability . The HPC community has studied performance variability throughout its history (e.g. [26], [27]), and has developed techniques towards understanding and managing this variability [28]. While some of their approaches to handle performance variance apply directly to resources in the cloud [29], [30], cloud-specific problems can require creative solutions that are applicable to a broad class of systems and applications. We focus on characterizing the variability of resources provided by some CSPs in an attempt to provide SMBs with a knowledge base to suit their application-specific needs.

1) Variability in the Cloud: Moving from dedicated infrastructure or services to the cloud imposes new challenges in solving the variability issues. The use of virtualized infrastructure in the cloud is likely to demonstrate higher load imbalance and performance fluctuations [31] due to factors such as resource sharing across multiple tenants and compute/memory resource over-subscription. Multiple research efforts ([32], [33]) have sought to provide effective performance isolation in multi-tenant data centers to minimize performance variability caused by multi-tenancy ("noisy neighbors") [34]. A challenging task for tenants is utilizing the large set of unpredictable variables to optimize their deployments and achieve a higher value for paid services. We hope to provide insight into the variability in the resource variable and enable them to make optimal choices in selecting a CSP.

2) Benchmarks: The CSPs in our study use a variety of hypervisor technologies to deliver their services. See table V for a summary. Evaluating their performance is difficult for several reasons. First, CSPs, like other technology vendors, may optimize their services to perform well on industry-standard benchmarks [35]. Second, the high degree of performance sensitivity to experimental design can potentially invalidate conclusions through the lack of repeatability and corroboration [36]. However, Saavedra and Smith [37] have shown that benchmarking the performance of various system components with a wide variety of micro-benchmarks and application kernels can provide a good estimate of that system's performance. We therefore choose to use a battery of micro-benchmark applications to evaluate the underlying subsystems, using the results to guide SMBs on the approximate performance to be expected on a CSP for the evaluated instances. We neglect the use of application kernels due to time constraints.

III. Design and Methodology

We enrolled in free trials on each CSP, and set up instances of varying sizes as allowed by the subscription. The instance types we used are: AWS: t2.micro; Azure: A0, A1, D1; Google: micro, n1.

CSP	(novice, intermediate, expert) participants
AWS	1, 2, 0
Google Cloud	1, 1, 2
Azure	1, 2, 3

TABLE III
Distribution of users in the usability study for each CSP.

A. Qualitative assessment

1) Usability: We evaluated the usability of IaaS interfaces using rigorous empirical evaluations [9], taking notes about user expressions and critical incidents [8]. We considered only the IaaS interface for each CSP for two reasons. First, inspection suggested that the interface designs for IaaS, PaaS, and SaaS are similar within a CSP, so this eliminated redundancy. Second, we felt that IaaS was conceptually the least complex because PaaS and SaaS layer on top of it. Focusing on the least complex option allowed us to open the study to intermediate users without concern that they might be overwhelmed, letting us observe the difference between the usability of each CSP's interface from both intermediate and expert perspectives.

An interface's effectiveness was evaluated through the participants' performances on a sequence of tasks (see table I). The tasks covered basic operations performed during the use of a CSP, requiring interaction with the administrative, technical, and billing areas of the interface. Researchers monitored each participant, measuring the Task Time to Complete (TTTC) and taking notes on any verbal or emotional expressions.

We enlisted participants with varying technical backgrounds and levels of domain knowledge. Participants were classified as intermediates or experts through a combination of self-reporting and their performance on a brief technical quiz (see table II). See table III for participant distributions across CSPs. Participants were not tested on CSPs with which they had previous experience.

2) Reliability: All the CSPs in our study offer SLAs with high levels of uptime, "or your money back" [16]–[18]. However, a refund is small potatoes compared to the loss of users or revenue a CSP customer might incur during an outage; the multi-day AWS outage in 2011 [38] serves as a vivid reminder of the potential impact of an outage.

To gauge the reliability of a CSP, we opted to measure the frequency of its **macro-outages**, those that had a significant effect on their customers. We define a macro-outage as one to which references appear in the first two pages of search results from a major search engine, reasoning that outages not significant enough to merit a news article or a blog post were not meaningful from the average customer's perspective. As the two major English-language search engines (Google and Bing) are operated by companies that also operate two of the three CSPs in our study, we searched for outages on each CSP using both of these search engines to avoid the risk of search favoritism [39]. We searched the following sequence of words: "*CSP* cloud outage *date*", using the values "AWS", "Azure", and "Google" for the "CSP" field, and choosing dates from "jan 2015" until "apr 2016". We ran these searches on

Number	Question	Type
1	Log into the instance management console.	Interface
2	How many instances are running?	Interface
3	What type(s) are they?	Technical
4	What are the CPU, memory, and OS of each instance?	Technical
5	Create an instance to host a database.	Technical
6	Try to ping the database instance.	Technical
7	Log into the database instance.	Technical
8	What was the last measured CPU utilization of the database instance?	Interface
9	Configure the database instance to automatically shut off if it uses more than 90% of the CPU for 5 consecutive minutes.	Interface
10	Stop the database instance.	Interface
11	On this account, what is the limit on the number of instances of the type you created?	Interface
12	If you shut down your account today, how much would you pay at the end of the monthly billing cycle?	Interface
13	When does the billing cycle end?	Interface

TABLE I
USABILITY TASKS.

Number	Question
1	How comfortable are you with computers and technology?
2	Have you ever used the command line before?
3	Do you know what ssh is?
4	Suppose you are creating a {database, web} server. Pick one of {high, low} for each of {CPU, memory}, with at most one high.

TABLE II
QUESTIONS USED TO GAUGE PARTICIPANT TECHNICAL EXPERTISE.

Score	Response characteristics
3	They fully answered the question.
2.5	They fully answered the question unprofessionally
2	You would need to follow up to get a complete answer.
1	You're more confused; they didn't even try.
0	They didn't respond.

TABLE IV
RUBRIC USED TO SCORE RESPONSES TO BILLING QUESTIONS.

	AWS	Google	Azure
Instance Type	t2.micro	N1	D1
Compute Unit	2 EC2CU	2.75 GCEU	160 ACU
Xeon Variant	E2670	E2670	E2660
Hypervisor	Xen	KVM	Hyper-V

TABLE V
COMPARISON OF CSP VCPUS WITH EQUIVALENT COMPUTE UNITS.

13 April 2016, covering a span of 15.4 months. Any outages occurring later in April were not considered. We did not differentiate between outages in IaaS, PaaS, SaaS, or any other CSP services.

We chose to only consider outages in 2015-2016 for two reasons. First, the historical outage rate is less relevant to a prospective CSP customer than the current outage rate. Second, it was clear from an inspection of AWS's annual outage frequency since inception that it takes some time to get a cloud service into a reliable state. Beginning in 2015 gave Azure and Google 2.5 years([40], [41]) to "practice".

3) Customer service: The only human contact available to users of the free tiers across all the CSPs we studied is email-based customer support for billing questions. Drawing on [25], we assessed the quality of their support by submitting 25 help tickets to each CSP, measuring the response time in hours and the quality of the response using the rubric in table IV. Our focus was on the responsiveness (response time) and reliability/assurance (response quality) components of SERVQUAL. As all of the CSPs we considered advertised responses within 2-3 business days, we discounted any hours that occurred on Saturdays and Sundays in EST. Any questions that did not receive a response within 5 business days were marked as "timed out".

B. Quantitative assessment

We used micro-benchmarks to compare the CSPs, evaluating the CPU, memory, and I/O performance of different instances. The results of our experiments can be used to estimate the relative performance of an application on instances from each CSP if it is strongly CPU-, memory-, or I/O-bound. We believe that the evaluations and benchmarks in this section provide necessary performance information to SMBs looking to make a transition to the cloud.

We deployed similar Linux instances on each CSP, taking benchmark measurements every 30 minutes for 2.5 days. This provides a high confidence interval and lets us evaluate the possibility of variation due to the time of the day. In addition to a set of homebrew microbenchmarks, we used the following industry-standard microbenchmarks: sysbench [42] and cpupi [43] (CPU), bandwidth64 [44] (memory), and sysbench and IOZone [45] (I/O).

Our evaluation of computational performance focuses on concretizing the virtual CPUs (vCPUs) each CSP offers. Table V shows how each CSP defines its basic unit of computation on the instances we studied. We used single-threaded CPU workloads to obtain a consistent comparison across instance types that have varying numbers of vCPUs (0.5, 1, or 2). Our home-brew CPU benchmarks use a tight loop of ALU, FPU, and NOOP operations to draw out any variability in processor cycles available to the instance. Turning to sysbench and cpupi allows us to measure vCPU performance on computationally intensive tasks. Combining these results lets us map out single-threaded performance for a vCPU processor speed and type.

Our evaluation of the memory subsystem is a measure of the bandwidth available to one core of the virtual machine. As proven by previous research [28], [46], [47], the memory subsystem can be a considerable bottleneck in a shared memory node. We stress the memory with a mix of read and write operations. Our memory benchmarks use up to 90% of all available memory, cycling through memory banks and overpopulating the cache to ensure cache misses, revealing any inconsistencies in the system memory throughput.

The I/O subsystem can be another key contributor to performance variability [48]–[51]. Each CSP advertises a multitude of storage offerings varying in configuration, complexity and scale, making it difficult to find a fair comparison. As the CSPs all offer directly-attached storage and provide QoS guarantees for it in terms of IOPs, we chose to measure the performance of this aspect of I/O. Our experiments used IOZone to generate a steady workload size of 10MB, with a record size of 32 KB and a cache overflow size of 15MB. The benchmark creates a total file set of 1 GB, with a non-uniform access distribution so that it has a cache miss ratio as high as 90%. During the execution of each workload, we monitor the performance of the following operations to evaluate the suitability of a CSP for a particular workload: random writes; sequential writes; sequential reads; re-reads; re-writes. Single-threaded and multi-threaded performance is measured to determine whether the benchmarks are reaching limits of the CPU or of the I/O subsystem.

IV. Findings and discussion

A. Qualitative Assessment

1) Usability Testing: We present findings from the usability studies conducted for each CSP. We use a combination of TTTC and observations and expressions made by participants to parse out empirical comparisons between the usability of each CSP. Microsoft Azure's interface stood out from the beginning as the most pleasant interface to interact with; 5 of the 6 participants who tested Azure expressed this verbally during session exit interviews. The emotional impact of using a system contributes to its overall usability, particularly to the user's willingness and ability to learn how to perform new operations on a system [52], [53].

Table VI shows the TTTC for technical tasks on Azure and Google for our two expert participants[1]. The expert participants took less time to finish technical tasks on Google Cloud than they did on Microsoft Azure. We believe that this is due to an interface design choice: the Google interface is flat, while Azure follows a hierarchical system of menus that give the user increasingly precise information or functionality. In contrast, there was no significant difference between the experts' results on Google Cloud or Azure on non-technical tasks. This could be because CSPs use standard terminology for technical concepts (e.g. *instance*, *cores*, *SSH*) but their own vocabulary for non-technical ones (e.g. Azure's *console*, Google Cloud's *dashboard*). If an expert recognizes the term he needs, he can jump right to it on a flatter interface, while the need to learn CSP-specific vocabulary might eliminate the benefit afforded by a flatter interface.

[1]We could not locate expert-level participants unfamiliar with AWS.

Task #	Expert #1 TTTC	Expert #2 TTTC
3	(124, 20)	(55, 30)
4	(104, 118)	(220, 95)
5	(135, 43)	(300, 54)
6	(18, 8)	(10, 13)
7	(20, 20)	(65, 22)

TABLE VI
TTTC in seconds for Expert Participants on technical tasks on (Azure, Google).

There were no other noticeable trends in the TTTC between Google and Azure. Sometimes, intermediate participants failed to complete a task within the allotted time, or gave up on a task of their own accord. We observed that of the three interfaces, AWS was the most prone to intermediate timeouts.

The observations above suggest that, whether intentionally designed this way or not, **Google's interface is aimed more toward expert users, while Azure is a little friendlier toward intermediate users.** We believe this is because the Google interface has a habit of providing a lot of information on a single page, while Azure's use of menus and submenus leads the user toward his desired operation.

There were particular tasks on which the participants' performances revealed **potential usability problems** in some of the interfaces. Two such tasks were task #7 and #4. On #7, we asked the user to log into an instance they had just created. Google's interface provided a convenient button labeled 'SSH' that opens an in-browser terminal emulator, already logged into the instance. Azure and AWS do not provide this, adding steps to the login process. On #4, we asked the user to find CPU, memory, and OS information about an instance that was already created. Information about the OS is not clearly available on any interface; the user would have to log into the instance and locate the file that contains that information.

2) Reliability: Table VII shows the macro-outage information for each CSP. AWS has the least outages and the smallest total outage time (about 16 hours over 15.4 months), and Azure and Google have more outages and 2x and 2.9x the total outage time. As businesses relying on an in-house email solution average 66-150 minutes of downtime per month [54], **AWS and Azure both compare favorably**.

3) Customer service: Table VIII summarizes the average response time in hours and response score for the 25 billing questions we submitted to each CSP. Each CSP averaged responses within its advertised 2-3 business day window. In addition, each observed that many of our questions were on related topics, and opted to close some as duplicates to address in a few larger tickets. Google's team was the most aggressive in this practice, closing 19 of the 25 tickets and responding on

CSP	Number of outages	(Total, average per outage, average per month) downtime	Total outage time relative to AWS
AWS [55]–[60]	5	(982, 196, 64)	1
Azure [61]–[69]	8	(1980, 247.5, 129)	2
Google [70]–[84]	12	(2883, 240.3, 187)	2.9

TABLE VII
MACRO-OUTAGES: 2015-2016 CSP OUTAGES MEASURED BY SEARCH ENGINE HITS.

the remaining 6. However, Google ultimately failed to respond to **9 of these 19 closed tickets** within 5 business days, yielding a response rate of only 64%. Amusingly, this is worse than the response rate of some hotels to email queries [85] in similar studies. Table VIII also indicates that AWS and Azure had comparable response quality, but that Azure's responses were 2.5x more rapid.

B. Quantitative Assessment

The goal in our experiments was to characterize the trade-offs between raw performance and variability in each CSP. To this end, we studied the variability of the CPU, memory, and I/O performance.

The box and whisker plot in figure 1 depicts the variability observed in all the CSPs for each of the tested characteristics. For CPU we show ALU performance, memory shows read-write speeds, and I/O shows random read-write performance. The variation in these benchmarks is representative across benchmarks for each resource. The range of variability has been mapped to Millions of OPs per second (MOP/s) for CPU, Gigabytes per second (GB/s) for memory, and Megabytes per second (MB/s) for file system I/O.

1) CPU: As seen in figure 1 (CPU) the variability observed in the vCPU performance is fairly consistent. Apart from a shared vCPU from Google's micro instance which consists of a shared vCPU, we see variability within a range of 5-10%. We believe that the impact of co-location of workloads adds considerably to the variability in performance. This implies that hardware virtualization assistance along with the CPU scheduling on the hypervisors of the respective instances are fairly consistent. Based on our empirical observations, we suggest the vCPU performance for dedicated vCPUs can enable compute intensive workloads with a reliable quality of service. We also observe that the performance of a single vCPU can vary by up to 5x both within (Azure A0 vs. Azure D1) and across a CSP (Google n1 vs. Azure A0). However, **there was no discrepancy between our results and the CSPs' internal performance metrics.** The performance ratio between and A0 and a A1 [86] is predicted to be 2x while we observed an average ratio of 1.998x. The performance on Google's shared and non-shared vCPU [87] was also as expected.

2) Memory: From figure 1 (Memory) we observe that the average memory write performance ranges across all CSPs in 2 distinct bands (1.5-3.1 GB/s and 10.3-21.1 GB/s). As demonstrated by [28], [88], [89], this amount of variability in memory throughput can significantly affect application performance. Azure D1 was an extreme case where the variability was observed with a standard variation exceeding 75% ($\pm$16%) of the average. Otherwise, variability was observed to be comparably low within the VM instances, but can still be as high as 15% ($\pm$8%) for Amazon EC2's t2.micro instance. Our results for Azure and Google indicate that **CSPs do not include memory throughput in their QoS guarantees**.

3) I/O: Figure 1 (I/O) illustrates the variability in the random read-write performance of the I/O subsystem. Each evaluation used dedicated attached storage. Azure D1 ran on an SSD and the others used HDDs. The average bandwidth recorded for random I/O ranged from 0.9 MB/s to 31.7 MB/s. The highest variability was observed on the Amazon AWS instance with a standard variation of 9% ($\pm$5%). Though all CSPs offer QoS guarantees in IOPs [90]–[92], Azure's definition of an IOP is unclear, which may explain the relatively slow performance of all of its instances. We note that these results disagree with [29], [48], which found significantly higher variability in I/O performance. This is likely due to our use of entry-level instances; **we hypothesize that the constraint on IOPs is limiting the variability in performance.**

As variability is present in all CSPs for different resources, SMBs should consider the requirements of their applications when choosing a CSP. Our vCPU measurements confirm the CSPs' internal performance metric, so SMBs should focus on memory and I/O variability as they consider an appropriate instance.

V. LESSONS LEARNED

We identified many areas for improvement and additional study in both the qualitative and quantitative portions of our work. Here are our favorites, in no particular order:

1) There are many more CSPs [1], and each bears examination.
2) Using Nielson's $N = 5$ recommendation [93] for both intermediate and expert usability study participants would have let us draw broader conclusions about usability.
3) Our usability interview process would have benefited from the more disciplined approach to interviewing taught by Rossman and Rallis [94], including obtaining and coding transcripts for each interview.
4) Comparing the responses of the billing and technical support staff of a CSP to identical queries raised by customers at each of the support tiers would indicate the varying skill of the support personnel at different tiers.
5) Each CSP offers a wide variety of instances. We were only able to study the least expensive instances due to our limited "free trial" funds.

CSP	Response rate	Average response time (hours)	Average response score
AWS	100%	7.5	2.7
Azure	100%	3	2.7
Google	64%	23.4	2.7

TABLE VIII

SUMMARY OF RESPONSES TO BILLING QUESTIONS. AVERAGE RESPONSE TIMES AND SCORES ONLY CONSIDER QUESTIONS THAT RECEIVED A RESPONSE.

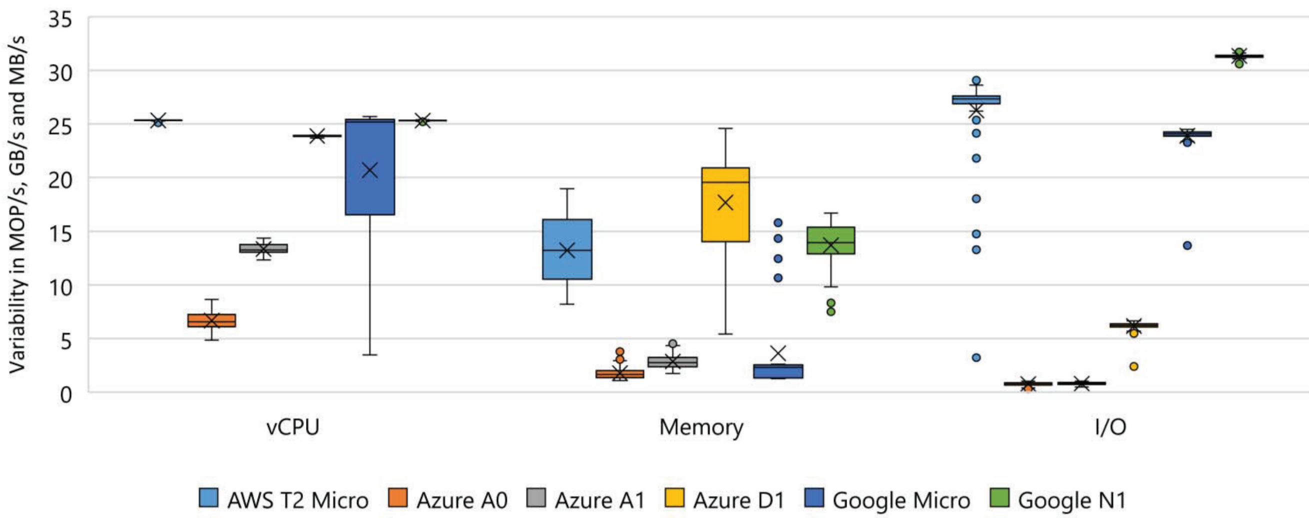

Fig. 1. The variability observed across all CSP instances for CPU, memory, and I/O.

6) It is not possible to draw application-level performance conclusions from micro-benchmark performance measurements. Cloud benchmarks relevant to applications deployed by SMBs would have made our findings more directly applicable.
7) Introducing multi-threaded tests would let us study the scale-up factor available to users of multi-CPU instances.
8) We took an *ad hoc* approach to dealing with the CPU bursting feature [95], [96] available on these CSPs, and we expect that this introduced some unnecessary variability into our CPU and memory studies.

VI. CONCLUSION

CSPs have changed the landscape for an SMB looking to make use of computational resources. No longer is an extensive on-site IT staff and infrastructure required. However, our study has found significant variation in the qualitative aspects of a CSP, as well as quantitative results that an SMB should take into consideration when choosing a CSP. For an SMB with weak technical skills or a small budget, our findings suggest the choice of Microsoft Azure among those CSPs we studied. 5 out of 6 users in our participant study stated that it had the most pleasant interface, its reliability is comparable to what an SMB would experience with an in-house solution, and it had the strongest customer service ratings. We believe that more performance-oriented firms can trust the CPU performance estimates provided by CSPs, and should focus their efforts on benchmarking memory and I/O performance as they investigate appropriate instance sizing for their use cases.

VII. ACKNOWLEDGMENTS

The authors would like to thank Kirsten, Chelsea, and US immigration services for their patience. We are grateful to our usability participants for their endurance, and to nameless customer support representatives for their unwitting contributions. Dr. Butt's whip-cracking drove us to greater heights, for which we are appreciative.

REFERENCES

[1] C. Staff. (2016, Feb.) The 100 coolest cloud computing vendors of 2016. [Online]. Available: http://www.crn.com/news/cloud/300079585/the-100-coolest-cloud-computing-vendors-of-2016.htm
[2] R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation computer systems*, vol. 25, no. 6, pp. 599–616, 2009.
[3] C. Staff. (2016, Apr.) Big four still dominate cloud infrastructures market in q1. [Online]. Available: http://www.crn.com/slide-shows/cloud/300079669/keeping-up-with-the-cloud-top-5-market-share-leaders.htm/pgno/0/6
[4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica *et al.*, "A view of cloud computing," *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.
[5] S. K. Garg, S. Versteeg, and R. Buyya, "A framework for ranking of cloud computing services," *Future Generation Computer Systems*, vol. 29, no. 4, pp. 1012–1023, 2013.

[6] X. Zheng, P. martin, K. Brohman, and L. Da Xu, “Cloudqual: a quality model for cloud services,” *Industrial Informatics, IEEE Transactions on*, vol. 10, no. 2, pp. 1527–1536, 2014.
[7] J. Nielsen and R. Molich, “Heuristic evaluation of user interfaces,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI ’90. New York, NY, USA: ACM, 1990, pp. 249–256. [Online]. Available: http://doi.acm.org/10.1145/97243.97281
[8] J. C. Flanagan, “The critical incident technique.” *Psychological bulletin*, vol. 51, no. 4, p. 327, 1954.
[9] J. Nielsen, “Usability inspection methods,” in *Conference companion on Human factors in computing systems*. ACM, 1994, pp. 413–414.
[10] P. Leitner, A. Michlmayr, F. Rosenberg, and S. Dustdar, “Monitoring, prediction and prevention of sla violations in composite services,” in *Web Services (ICWS), 2010 IEEE International Conference on*. IEEE, 2010, pp. 369–376.
[11] V. C. Emeakaroha, m. A. Netto, R. N. Calheiros, I. Brandic, R. Buyya, and C. A. De Rose, “Towards autonomic detection of sla violations in cloud infrastructures,” *Future Generation Computer Systems*, vol. 28, no. 7, pp. 1017–1029, 2012.
[12] H. Zhang, L. Ye, J. Shi, X. Du, and M. Guizani, “Verifying cloud service-level agreement by a third-party auditor,” *Security and Communication Networks*, vol. 7, no. 3, pp. 492–502, 2014.
[13] CloudHarmony. (2016, Jan.) Research and compare cloud providers and services. [Online]. Available: https://cloudharmony.com/status-1year-group-by-regions-and-provider
[14] S. A. Baset, “Cloud slas: present and future,” *ACM SIGOPS Operating Systems Review*, vol. 46, no. 2, pp. 57–66, 2012.
[15] J.-H. Morin, J. Aubert, and B. Gateau, “Towards cloud computing sla risk management: issues and challenges,” in *System Science (HICSS), 2012 45th Hawaii International Conference on*. IEEE, 2012, pp. 5509–5514.
[16] Amazon. (2013, Jun.) Amazon ec2 service level agreement. [Online]. Available: http://aws.amazon.com/ec2/sla/
[17] Microsoft. (2016, Apr.) Sla for virtual machines. [Online]. Available: https://azure.microsoft.com/en-us/support/legal/sla/virtual-machines/v1_1/
[18] Google. (2016, Apr.) Google compute engine service level agreement (sla). [Online]. Available: https://cloud.google.com/compute/sla
[19] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, “Servqual,” *Journal of retailing*, vol. 64, no. 1, pp. 12–40, 1988.
[20] R. N. Bolton and J. H. Drew, “A longitudinal analysis of the impact of service changes on customer attitudes,” *The Journal of marketing*, pp. 1–9, 1991.
[21] J. J. Cronin Jr and S. A. Taylor, “Measuring service quality: a reexamination and extension,” *The journal of marketing*, pp. 55–68, 1992.
[22] S. J. Barnes and R. T. Vidgen, “An integrative approach to the assessment of e-commerce quality.” *J. Electron. Commerce Res.*, vol. 3, no. 3, pp. 114–127, 2002.
[23] S. Han and S. Baek, “Antecedents and consequences of service quality in online banking: An application of the servqual instrument,” *Advances in Consumer Research*, vol. 31, no. 2, pp. 208–14, 2004.
[24] A. Elmorshidy, “Applying the technology acceptance and service quality models to live customer support chat for e-commerce websites,” *Journal of Applied Business Research (JABR)*, vol. 29, no. 2, pp. 589–596, 2013.
[25] J. Murphy, R. Schegg, and D. Olaru, “Quality clusters: Dimensions of email responses by luxury hotels,” *International Journal of Hospitality Management*, vol. 26, no. 3, pp. 743–747, 2007.
[26] G. Lasseter, T. Lo, K. M. Chandy, and J. C. Browne, *Statistical and Pattern Based Models for CPU Burst Prediction*. Computer Science Department, University of Texas at Austin, 1973.
[27] R. Mraz, “Reducing the variance of point to point transfers in the ibm 9076 parallel computer,” in *Proceedings of the 1994 ACM/IEEE conference on Supercomputing*. IEEE Computer Society Press, 1994, pp. 620–629.
[28] D. Skinner and W. Kramer, “Understanding the causes of performance variability in hpc workloads,” in *Workload Characterization Symposium, 2005. Proceedings of the IEEE International*. IEEE, 2005, pp. 137–149.
[29] A. Iosup, N. Yigitbasi, and D. Epema, “On the performance variability of production cloud services,” in *Cluster, Cloud and Grid Computing (CCGrid), 2011 11th IEEE/ACM International Symposium on*. IEEE, 2011, pp. 104–113.
[30] J. Schad, J. Dittrich, and J.-A. Quiané-Ruiz, “Runtime measurements in the cloud: observing, analyzing, and reducing variance,” *Proceedings of the VLDB Endowment*, vol. 3, no. 1-2, pp. 460–471, 2010.
[31] Y. Cheng, A. Gupta, and A. R. Butt, “An in-memory object caching framework with adaptive load balancing,” in *Proceedings of the Tenth European Conference on Computer Systems*, ser. EuroSys ’15. New York, NY, USA: ACM, 2015, pp. 4:1–4:16. [Online]. Available: http://doi.acm.org/10.1145/2741948.2741967
[32] D. Shue, M. J. Freedman, and A. Shaikh, “Performance isolation and fairness for multi-tenant cloud storage.” in *OSDI*, vol. 12, 2012, pp. 349–362.
[33] R. Kapoor, G. Porter, M. Tewari, G. M. Voelker, and A. Vahdat, “Chronos: Predictable low latency for data center applications,” in *Proceedings of the Third ACM Symposium on Cloud Computing*, ser. SoCC ’12. New York, NY, USA: ACM, 2012, pp. 9:1–9:14. [Online]. Available: http://doi.acm.org/10.1145/2391229.2391238
[34] S. Angel, H. Ballani, T. Karagiannis, G. O’Shea, and E. Thereska, “End-to-end performance isolation through virtual datacenters,” in *11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14)*. Broomfield, CO: USENIX Association, Oct. 2014, pp. 233–248. [Online]. Available: https://www.usenix.org/conference/osdi14/technical-sessions/presentation/angel
[35] E. Mack, “HTC admits boosting One M8 benchmarks; makes it a feature,” 2014, http://www.cnet.com/news/is-the-htc-one-m8-that-good-benchmark-cheating-alleged-again/.
[36] J. Vitek and T. Kalibera, “Repeatability, reproducibility, and rigor in systems research,” in *Proceedings of the ninth ACM international conference on Embedded software*. ACM, 2011, pp. 33–38.
[37] R. H. Saavedra and A. J. Smith, “Analysis of benchmark characteristics and benchmark performance prediction,” *ACM Trans. Comput. Syst.*, vol. 14, no. 4, pp. 344–384, Nov. 1996. [Online]. Available: http://doi.acm.org/10.1145/235543.235545
[38] AWS. (2011, Apr.) Summary of the amazon ec2 and amazon rds service disruption in the us east region. [Online]. Available: http://aws.amazon.com/message/65648/
[39] I. Kottasova and D. Goldman. (2015, Apr.) Google under siege: Europe wants blood. [Online]. Available: http://money.cnn.com/2015/04/15/technology/google-europe-anti-trust-lawsuit/index.html
[40] C. McLuckie. (2012, Jun.) Google compute engine: Computing without limits. [Online]. Available: http://googleforwork.blogspot.com/2012/06/google-compute-engine-computing-without.html?utm_source=entblog&utm_medium=blog&utm_campaign=Feed:+OfficialGoogleEnterpriseBlog+%28Official+Google+Enterprise+Blog%29
[41] B. Laing. (2012, Jun.) Announcing new windows azure services to deliver “hybrid cloud”. [Online]. Available: https://azure.microsoft.com/en-us/blog/announcing-new-windows-azure-services-to-deliver-hybrid-cloud/
[42] A. Kopytov, “Sysbench: a system performance benchmark,” *URL: http://sysbench. sourceforge. net*, 2004.
[43] M. Tommila. (2005, Feb.) apfloat: A c++ high performance arbitrary precision arithmetic package. [Online]. Available: http://www.apfloat.org/apfloat/
[44] Z. Smith. (2016) Bandwidth: a memory bandwidth benchmark. [Online]. Available: http://zsmith.co/bandwidth.html
[45] W. D. Norcott and D. Capps, “Iozone filesystem benchmark,” *URL: www. iozone. org*, vol. 55, 2003.
[46] N. R. Mahapatra and B. Venkatrao, “The processor-memory bottleneck: problems and solutions,” *Crossroads*, vol. 5, no. 3es, p. 2, 1999.
[47] D. Xu, C. Wu, and P.-C. Yew, “On mitigating memory bandwidth contention through bandwidth-aware scheduling,” in *Proceedings of the 19th international conference on Parallel architectures and compilation techniques*. ACM, 2010, pp. 237–248.
[48] D. Ghoshal, R. S. Canon, and L. Ramakrishnan, “I/o performance of virtualized cloud environments,” in *Proceedings of the second international workshop on Data intensive computing in the clouds*. ACM, 2011, pp. 71–80.
[49] J. Borrill, L. Oliker, J. Shalf, and H. Shan, “Investigation of leading hpc i/o performance using a scientific-application derived benchmark,” in *Proceedings of the 2007 ACM/IEEE conference on Supercomputing*. ACM, 2007, p. 10.
[50] M. Liu, J. Zhai, Y. Zhai, X. Ma, and W. Chen, “One optimized i/o configuration per hpc application: leveraging the configurability of cloud,” in *Proceedings of the Second Asia-Pacific Workshop on Systems*. ACM, 2011, p. 15.

[51] S. Lang, P. Carns, R. Latham, R. Ross, K. Harms, and W. Allcock, "I/o performance challenges at leadership scale," in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*. ACM, 2009, p. 40.
[52] D. Norman, "Emotion & design: Attractive things work better," *interactions*, vol. 9, no. 4, pp. 36–42, Jul. 2002. [Online]. Available: http://doi.acm.org/10.1145/543434.543435
[53] K. Boehner, R. DePaula, P. Dourish, and P. Sengers, "Affect: From information to interaction," in *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility*, ser. CC '05. New York, NY, USA: ACM, 2005, pp. 59–68. [Online]. Available: http://doi.acm.org/10.1145/1094562.1094570
[54] Google, "What we learned from 1 million businesses in the cloud," 2008, https://googleblog.blogspot.com/2008/10/what-we-learned-from-1-million.html.
[55] J. Ryall. (2016, Mar.) Telstra's problems in australia continue with amazon web services outage. [Online]. Available: http://mashable.com/2016/03/17/telstra-outage-affects-aws/#vpx3zo3IM5qD
[56] N. Heath. (2015, Sep.) Aws outage: How netflix weathered the storm by preparing for the worst. [Online]. Available: http://techrepublic.com/article/aws_outage-how-netflix-weathered-the-storm-by-preparing-for-the-worst/
[57] M. Smolaks. (2015, Sep.) Aws suffers a five-hour outage in the us. [Online]. Available: http://www.datacenterdynamics.com/colo-cloud-/aws-suffers-a-five-hour-outage-in-the-us/94841.fullarticle
[58] J. Tsidulko. (2015, Aug.) Overnight aws outage reminds world how important aws stability really is. [Online]. Available: http://www.crn.com/news/cloud/300077740/overnight-aws-outage-reminds-world-how-important-aws-stability-real_ly-is.htm
[59] O. Williams. (2015, Jul.) Favorite service down? amazon web services is having issues. [Online]. Available: http://thenextweb.com/insider/2015/07/31/favorite-service-down-amazon-web-services-is-having-issues/#gref
[60] N. Kephart. (2015, Jun.) Route leak causes amazon and aws outage. [Online]. Available: https://blog.thousandeyes.com/route-leak-causes-amazon-and-aws-outage/
[61] P. Gothard. (2016, Feb.) Office 365 suffers european outage due to 'high resource utilisation'. [Online]. Available: http://www.computing.co.uk/ctg/news/2447946/office-365-suffers-global-outage-due-to-high-resource-utilisation
[62] ——. (2015, Dec.) Another microsoft office 365 and azure outage hits uk and europe. [Online]. Available: http://www.computing.co.uk/ctg/news/2439832/another-microsoft-office-365-and-azure-outage-hits-uk-and-europe
[63] ——. (2015, Dec.) Microsoft office 365 and azure experience widespread outage. [Online]. Available: http://www.computing.co.uk/ctg/news/2437639/microsoft-office-365-experiences-widespread-outage
[64] C. Donnelly. (2015, Dec.) European office 365 and microsoft azure users hit by service outage. [Online]. Available: http://www.computerweekly.com/news/4500259905/European-Office-365-and-Microsoft-Azure-users-hit-by-service-outage
[65] B. Sullivan. (2015, Jul.) Microsoft hit with office 365 cloud outage. [Online]. Available: http://www.techweekeurope.co.uk/cloud/cloud-management/microsoft-cloud-outage-office-365-172673
[66] S. Sharwood. (2015, Jul.) Azure australia went titsup for about seven hours. [Online]. Available: http://m.theregister.co.uk/2015/07/03/azure_australia_in_titsup_hiccup/
[67] J. Saarinen. (2015, Jul.) Microsoft azure australia east hit by outage. [Online]. Available: http://www.itnews.com.au/news/microsoft-azure-australia-east-hit-by-outage-406110
[68] K. McLaughlin. (2015, May) Azure nightmare: Customer suffers 9-day intermittent outage, gets no help from microsoft. [Online]. Available: http://www.crn.com/news/cloud/300076751/azure-nightmare-customer-suffers-9-day-intermittent-outage-gets-no-h_elp-from-microsoft.htm
[69] J. Tsidulko. (2015, Mar.) Microsoft gets another headache: Azure outage redux. [Online]. Available: http://www.crn.com/news/cloud/300076173/microsoft-gets-another-headache-azure-outage-redux.htm
[70] S. Sharwood. (2016, Apr.) Google cloud titsup. [Online]. Available: http://www.theregister.co.uk/2016/04/12/google_cloud_titsup/
[71] JeffCraig. (2016, Apr.) Google cloud outage - sc, discord and more affected. [Online]. Available: https://testsquadron.com/threads/google-cloud-outage-sc-discord-and-more-affected.7135
[72] Google. (2016, Apr.) Google compute engine incident #16007. [Online]. Available: https://status.cloud.google.com/incident/compute/16007
[73] S. Perez. (2015, Dec.) Google app engine issues led to snapchat's extended outage. [Online]. Available: http://techcrunch.com/2015/12/08/snapchat-is-experiencing-an-extended-outage/
[74] E. Wingfield. (2015, Dec.) Recent google cloud outage attributed to case sensitive code deployment. [Online]. Available: http://www.cloudwedge.com/recent-google-cloud-outage-attributed-to-case-sensitive-code-de_ployment-8378/
[75] A. MacGregor. (2015, Nov.) Google suffers european cloud outage, human error to blame. [Online]. Available: https://thestack.com/cloud/2015/11/30/google-suffers-european-cloud-outage-human-error-to-blame/
[76] M. Wheatley. (2015, Nov.) Google cloud suffers new outage in europe. [Online]. Available: http://siliconangle.com/blog/2015/11/30/google-cloud-suffers-new-outage-in-europe/
[77] K. Bell. (2015, Oct.) Google drive is back online after brief outage. [Online]. Available: http://mashable.com/2015/10/09/google-docs-sheets-down/#KmBta_tG2aqR
[78] S. Sharwood. (2015, Aug.) Google cloud in nine-hour storage snafu. [Online]. Available: http://www.theregister.co.uk/2015/08/27/google_cloud_in_ninehour_storage_snafu/
[79] Google. (2015, Aug.) Google compute engine incident #15056. [Online]. Available: https://status.cloud.google.com/incident/compute/15056
[80] ——. (2015, Jun.) Google cloud storage incident #16025. [Online]. Available: https://status.cloud.google.com/incident/storage/16025
[81] D. Weldon. (2015, Apr.) Gmail outage traced to certificate expiration. [Online]. Available: http://www.fiercecio.com/story/gmail-outage-traced-certificate-expiration/2015-04-06
[82] J. Verge. (2015, Mar.) Google traces sundays cloud outage to faulty patch. [Online]. Available: http://www.datacenterknowledge.com/archives/2015/03/09/google-traces-sundays-cloud-outage-to-faulty-patch/
[83] S. Sharwood. (2015, Mar.) Google promises proper patch preparation after new cloud outage. [Online]. Available: http://www.theregister.co.uk/2015/03/09/google_promises_proper_patch_preparation_after_second_cloud_outage/
[84] J. Verge. (2015, Feb.) Google compute engine, aol mail suffer early morning outages. [Online]. Available: http://www.datacenterknowledge.com/archives/2015/02/19/google-suffers-cloud-outage-aol-mail-goes-down-early-morning/
[85] R. Schegg, A. Liebrich, F. Liu, and J. Murphy, "eservice by swiss and austrian hotels: does language matter?" *Information and communication technologies in tourism 2006*, pp. 357–368, 2006.
[86] Microsoft. Pricing - virtual machines (vms). [Online]. Available: https://azure.microsoft.com/en-us/pricing/details/virtual-machines/
[87] Google. Google compute engine pricing. [Online]. Available: https://cloud.google.com/compute/pricing
[88] F. Liu, X. Jiang, and Y. Solihin, "Understanding how off-chip memory bandwidth partitioning in chip multiprocessors affects system performance," in *High Performance Computer Architecture (HPCA), 2010 IEEE 16th International Symposium on*. IEEE, 2010, pp. 1–12.
[89] B. M. Rogers, A. Krishna, G. B. Bell, K. Vu, X. Jiang, and Y. Solihin, "Scaling the bandwidth wall: challenges in and avenues for cmp scaling," in *ACM SIGARCH Computer Architecture News*, vol. 37, no. 3. ACM, 2009, pp. 371–382.
[90] Google. (2016, May) Google compute iops service limits. [Online]. Available: https://cloud.google.com/compute/docs/disks/performance
[91] Microsoft. (2016, May) Azure cloud iops service limits. [Online]. Available: https://azure.microsoft.com/en-us/documentation/articles/azure-subscription-service-limits/
[92] Amazon. (2016, May) Elastic compute iops service limits. [Online]. Available: http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/ebs-io-characteristics.html
[93] J. Nielsen. (2012, Jan) Usability 101: Introduction to usability. [Online]. Available: https://www.nngroup.com/articles/usability-101-introduction-to-usability/
[94] G. Rossman and S. F. Rallis, "Learning in the field: An introduction to qualitative research," 2012.
[95] Amazon. Ec2 instance types. [Online]. Available: https://aws.amazon.com/ec2/instance-types/
[96] Google. (2016, April) Google compute engine pricing. [Online]. Available: https://cloud.google.com/compute/pricing