ABSTRACT

The original benchmark was the result of the collaboration between Burstorm and Rice University and uses a high degree of automation. The scope of the first benchmark was seven suppliers across three continents with a total of 96 different instance types. The scope of this report shows an increased number of instance types to 153, each tested in 3 locations for a total of 459 instances tested per cycle. Since June 2015 we have tested all available instances at random days and times several times a week (to cover any day of time and any day of the week) and are continuing to do so while adding new instance types. Today this represents about 23,000 data points. This updated version follows the same structure where possible.
Table of Contents

INTRODUCTION.............................................................................................................................................. 3
METHODOLOGY................................................................................................................................................ 6
  PROCESS......................................................................................................................................................... 7
  UPDATED SCOPE...................................................................................................................................... 9
BENCHMARK RESULTS .................................................................................................................................. 11
  SUMMARY ................................................................................................................................................... 11
  DEFINITIONS ........................................................................................................................................... 14
    Performance ........................................................................................................................................... 14
    CPU performance .................................................................................................................................. 14
    IO performance ..................................................................................................................................... 14
    Price ....................................................................................................................................................... 14
    Price-Performance (Updated) ................................................................................................................... 14
  PERFORMANCE BY CLOUD SERVICE PROVIDER .................................................................................. 15
    Absolute Performance .......................................................................................................................... 20
  PRICE PERFORMANCE ........................................................................................................................... 21
  PERFORMANCE OVER TIME .................................................................................................................... 25
  GLOBAL OBSERVATIONS .......................................................................................................................... 31
CONCLUSIONS .............................................................................................................................................. 37

APPENDIX 1: TEST DETAILS ....................................................................................................................... 38
  TEST DETAILS .......................................................................................................................................... 38
  BCU: SYSTEM SPECS ............................................................................................................................... 38

APPENDIX 2: WHAT’S NEXT ......................................................................................................................... 39

REVIEW HISTORY – V1 ............................................................................................................................... 40
Introduction

Consumer Internet businesses, like eBay, Twitter and Facebook depend on their computing infrastructure (compute, storage, data centers and networks) as the foundation of their enterprise. Increasingly this is true across other industries including high tech, financial services, biotech, healthcare, etc. These infrastructure components are more and more consumed as a service (Cloud computing).

Given the increasing complexity of cloud deployments, Burstorm in 2015 launched the industry’s first Computer-Aided Design (CAD) application for cloud architects. Like Autodesk in construction, Burstorm’s application allows architects to develop new infrastructure designs as well as remodel existing compute, storage datacenter and network infrastructures. The cornerstone of the application is a product catalog, which as of the writing of V1 contained over 900 product sets totaling over 36000 products. The product catalog today contains product specifications, pricing covering different types of business models and location information. Based on this product catalog and a class of optimization algorithms the application aids the architect in making design decisions.

Over the past several years, Dr. T. S. Eugene Ng’s group at Rice University has also been focused on cloud computing. One of the areas of research interest has been the performance of compute and storage cloud services. Recently they published their joint work with Purdue University: Application-Specific Configuration Selection in the Cloud: Impact of Provider Policy and Potential of Systematic Testing¹, in the Proceedings of IEEE INFOCOM’15. The paper takes a first step towards understanding the impact of cloud service provider policy and tackling the complexity of selecting configurations that can best meet the price and performance requirements of applications.

Their work sparked the interest of Edward Wustenhoff at Burstorm. At the same time, Dr. Ng was hoping to collaborate with practitioners to get exposed to a wider set of configuration choices, and other compute & storage cloud service providers, beyond Amazon EC2. There are a number of challenges to price-performance benchmarking since Jim Gray’s landmark paper, A Measure of Transaction Processing Power². First, as Burstorm’s product catalog shows there are now 1000s of different compute & storage cloud services. These cloud services span many different locations. One might think it’s odd to talk about location and cloud services in the same sentence, but for geopolitical and networking performance reasons the location of these cloud services does matter. Furthermore, there are a variety of business models. One can see several cloud price

² Jim Gray, A Measure of Transaction Processing Power, 1985
changes in a 24-month period. And, as we have observed, the performance of the same instance can be different at different times and different locations.

On top of that you can consume services by the hour, month, annually or one can buy on the spot market. New products are being introduced on a monthly basis and pricing can change weekly. For instance, Amazon made tens of price changes in a 24-month period. And as the INFOCOM’15 study has observed, the performance of the same instance can be different at different times and different locations.

The result of the collaboration with Rice was the industry’s first comprehensive and continuous price-performance benchmark. Using a high degree of automation the scope of the first benchmark was seven suppliers (Amazon, Google, Microsoft, Rackspace, IBM, HP and Linode), across three continents (Asia, North America and Europe) with a total of 266 compute products spread over 3 locations per vendor, where available. The benchmark was executed every day, for 15 days. The scope of V2 shows an increased number of instance types to 153, each tested in 3 locations for a total of 459 instances tested per cycle. Since June 2015 we have tested all available instances at random days and times several times a week (to cover any day of time and any day of the week) and are continuing to do so while adding new instance types. Today this represents more than 23,000 data points. The results are normalized to a 720-hour, monthly pricing model to establish the price-performance metrics.

Most of us are familiar with traditional performance testing. However we believe that those practices are only partially applicable to understanding cloud computing performance. What makes this report unique and interesting is that we tested a large amount of instance types (153) over time, in multiple locations and include economic impact data. Some of the results show a large variation of performance within a same instance type. The best performing instance does not show the best price-performance. Availability and behavior of instances is not the same depending on location within the same provider. All in all, the cloud is a very dynamic and complex environment.

This updated PDF report shows selected screen shots from the interactive report, which is available as part of the Burstorm Application. The interactive report allows you to visualize the data in many different ways and allowed us to create the updated information in this report. For example HP pulled out of the cloud market since the original report, Rackspace and Microsoft’s instance types increased significantly and we added Digital Ocean.
Our plans for the future include: adding more cloud service providers and locations, and the development of RB-B. More forward looking statements can be read in Appendix 2: What’s next.

As promised in the original report, Burstorm has recently incorporated the performance data into its CAD application so cloud architects can create architectures optimized for locality, price, performance and now price-performance. Burstorm also expanded the application by enabling benchmarking for dedicated and private cloud services. If you have any questions please contact Edward Wustenhoff.
Methodology

When we started thinking about what was different between the new performance dynamics in cloud computing and the TPC Benchmark days, we realized that because there is less control over the environment we cannot assume that every instance tested is identical at start up, over time and per location. This causes great uncertainty about the capability to process workloads consistently. In addition, all the new business models raise questions about the economic benefits for certain instance types. Selecting the optimal instance for a specific workload has become also a function of performance and economics.

We came to the realization that the only conclusive way to address this, would be by continually testing all instance types everywhere. One can imagine how this becomes a logistical and economic challenge that is seemingly impossible to address.

It’s because of this we came to change how we think of the benchmarking process and not so much about creating a better benchmark. Of course certain aspects within a Virtual Machine will need to be tested differently and Burstorm is working with Rice University on improving the benchmarks, but in the end the biggest challenge is around scale and velocity.

Fortunately, in the new compute era, the time and cost to create a benchmark environment can be measured in cents and minutes and is easily distributed through automation. The process and scope we applied are outlined below.
Process

Figure 1: RB-A Test Process shows the high-level process we use to spin up, benchmark, write results and display the results.

The basic concept is to spin up instances, run the benchmark, write the data to the Burstorm product catalog, combine it with our pricing data and continually repeat this several times a week at random days and times, for each instance type and for each provider at each selected locations.

Because not all providers in the target set have services in the same locations we decided to select one for each provider in Asia, the US and Europe so we could spot potential differences in deployment and cost per region. We intend to expand providers and locations as we continue our benchmarking.

It was interesting to experience the difference in deployment processes for each supplier. Some contacted us to ensure we were legitimate, others wanted financial guarantees and one would send us several emails for each instance type started to confirm approval, actual provisioning and a “Getting started” email. We also found some bugs in provider’s
deployment API’s we had to fix before we could proceed. This has proven to be a continual process.

Some required us to open up a separate account for other countries, all interesting indicators of the maturity level of this market place.

Benchmarks were run in parallel, though with some damping to avoid limits (cpu, memory etc.) of various providers. The instances were created using standard chef knife CLI commands (e.g. "knife [provider] server create"), which started and loaded the benchmark software onto the instance. When finished, the software reported back the test results to our server using a JSON version of the standard UnixBench test results.

Due to the scale and need for automation, we used best effort to gather the data for each test run, and as such there are sometimes missing data points. We allowed for this as opposed to trying to fix failures to launch because it is another interesting data point. However, since this is out of scope for this report, we haven't diagnosed deeply why the instances we tried to spin up didn't start, but some portion of it seems to hint towards capacity limitations on the provider's side because missing instance runs were often in the larger 8-16 core variety. At the time of writing the V2 report we still see this pattern as a common occurrence.

Burstorm uses the standard UnixBench score but scaled to a more modern processor and bus (a Raspberry Pi 2, ARM7 @900Mhz) from the original SparcStation 20-61 "George". The detailed specifications of the system and tests can be found in Appendix 1: Test details

This latest updated was to create the same views of the new data, the main content of this report. We combined the performance data with the product pricing catalog data from Burstorm's CAD application to create the price-performance benchmark numbers. This performance data is now also available in our CAD application and has become part of the design information of compute, storage, data center and network architectures.
**Updated Scope**

The below table shows the scope of this project creating the test results each with multiple data points. See Appendix 1: Test details for how the data points were created. Note that we confined the number of locations to three for the reasons mentioned earlier.

For this report we did not yet test any dedicated or “bare metal” instance types. We are currently working with several providers and these will be added in the future.

The original report included HP but since they stopped providing public cloud offerings in January 2016 we replaced them with Digital ocean.

<table>
<thead>
<tr>
<th>Provider</th>
<th># Instance Types</th>
<th># Locations</th>
<th># Products</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS</td>
<td>39 (+ 9)</td>
<td>3</td>
<td>117</td>
</tr>
<tr>
<td>Google</td>
<td>18 (+ 4)</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>Rackspace</td>
<td>25 (+ 16)</td>
<td>3</td>
<td>75</td>
</tr>
<tr>
<td>Azure</td>
<td>39 (+ 21)</td>
<td>3</td>
<td>117</td>
</tr>
<tr>
<td>Linode</td>
<td>9 (+ 0)</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td><strong>HP</strong></td>
<td><strong>11</strong> (Deleted)</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Digital Ocean</td>
<td>18 (new)</td>
<td>3</td>
<td>54</td>
</tr>
<tr>
<td>Softlayer</td>
<td>5 (+ 0)</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>

**Selected total** | **153 (+ 57)** | **21 (+ 2)** | **459 (+ 193)**

*Table 1: Testing Scope*

The following locations were selected for each region by provider:

<table>
<thead>
<tr>
<th>Provider</th>
<th>North America (NA)</th>
<th>Europe (EMEA)</th>
<th>Asia (APAC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS</td>
<td>Ashburn US</td>
<td>Dublin IE</td>
<td>Singapore SG</td>
</tr>
<tr>
<td>Google</td>
<td>Council Bluffs US</td>
<td>Saint-Ghislain BE</td>
<td>Changhua County TW</td>
</tr>
<tr>
<td>Rackspace</td>
<td>Grapevine US</td>
<td>Slough GB</td>
<td>Hong Kong HK</td>
</tr>
<tr>
<td>Azure</td>
<td>California, CA</td>
<td>Omeath IE</td>
<td>Singapore SG</td>
</tr>
<tr>
<td>Linode</td>
<td>Fremont US</td>
<td>London GB</td>
<td>Singapore SG</td>
</tr>
<tr>
<td>HP</td>
<td>Tulsa US</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Digital Ocean</td>
<td>San Francisco</td>
<td>Amsterdam NL</td>
<td>Singapore SG</td>
</tr>
<tr>
<td>Softlayer</td>
<td>San Jose US</td>
<td>Amsterdam NL</td>
<td>Singapore SG</td>
</tr>
</tbody>
</table>

*Table 2: Locations by provider*
We did not separately test Windows instances for the following reasons:

- Not all providers have windows instances and we wanted to make sure we had a common baseline.
- Assuming the impact of the underlying virtualization to be equal for any OS, we expect the relative performance between 4core and 8core systems to be somewhat equal for both Windows and Linux.

We are pursuing an equivalent test for Windows and possibly other operating systems.

This report reflects an updated version of the 1st comprehensive price-performance benchmark published in June 2015. The following results are just one view of the data, which can be analyzed in many other ways in the interactive report.
Benchmark Results

Summary

This update (V2) to the first comprehensive and continuous price-performance benchmark has yielded some interesting observations:

- Performance of 1-core instances can still vary by 615% between providers.

![Figure 2: Performance scores for 1 core instances](image)

- MSFT now has the top performer spot with their G5 instance, followed by AWS’s m4.10xlarge and c3.8xlarge.

![Figure 3: MSFT G5 test results](image)
• **Price performance for a 4-core compute cloud service can vary by 1501%**
  The top three price-performance winners for 4 core systems were Linode-4GB, Digital Ocean’s 8GB instance and Rackspace’s General 1-4.

![Figure 4: Price/Performance of 4 Core instances](image)

• **The same instance performance can still fluctuate by 62% over time.**
  However, it seems that size matters in 2 ways: More performance volatility seem to be more prevalent at the larger instance types and the larger providers like AWS, Google, Rackspace and MSFT seem to provide more consistent performance over time than the smaller providers.

![Figure 5: Performance over time, 4Core, Digital Ocean instances (Max=18.95, Min=11.69)](image)
Not all locations are created equal in availability and performance of instance types. Most noticeable is that not all instances are available everywhere. Not all high performance Microsoft Azure instance types are available in APAC for example and the G4 & 5, which are new, seemed not yet available everywhere either. However, as a general observation, consistency seems to improve. Comparing 8 Core compute cloud services of Google Compute Engine between current (V2) and last (V1) report shows considerably less difference.

The rate of change in instance types, pricing, performance over time and availability of services by location confirms that the traditional way of benchmarking a small set of instance types in a unique event is not sufficient anymore in today’s world of cloud computing. To see the longer term trends and understand the wide variety of results we created the interactive report for our customers.

Continuous and comprehensive benchmarking of existing and new cloud services will reveal useful information for both suppliers and consumers of compute & storage cloud services. Rice and Burstorm will continue to expand the scope of the benchmarking and work with enterprises, academics and cloud service providers to add to our collective understanding of the cloud. If you have any questions please contact Edward Wustenhoff.

The next chapters show the details and data that led up to these findings. But let’s provide you with some definition of terms first.
Definitions
In order to better understand the graphs and statements made, below are the definitions of the key metrics used in this and the interactive report.

Performance
This reflects the UnixBench score relative to the Burststorm Compute Unit (BCU) baseline. See more on that in Appendix 1: Test details. When multiple data points applied an average of the scores was taken. A higher score means better performance.

CPU performance
CPU performance is measured using a subset of the UnixBench tests, namely:
1. dhry2reg -- Dhrystone CPU using two register variables
2. whetstone-double -- Whetstone double precision CPU test
3. pipe -- Unix pipe throughput
4. context1 -- Pipe based context switching throughput
5. shell8 -- 8 bash shells executing simultaneously

A higher score means better CPU performance.

IO performance
IO performance is measured using a subset of the UnixBench tests, namely:
1. fstime -- file copy, 1024 byte buffer size, 500 maxblocks
2. fsbuffer -- file copy, 256 byte buffer size, 500 maxblocks
3. fsdisk -- file copy, 4096 by buffer size, 8000 maxblocks

A higher score means better IO performance.

Price
Price is the monthly cost using hour-hour terms, normalized to 720/hrs/month, no prepayments and using Ubuntu 14.04 Linux. The prices used in this document reflect the prices of the instance running the specified OS at the start of the test period. Realize that a Redhat or Windows OS instance types would typically carry a higher price.

Price-Performance (Updated)
Price performance is defined as the monthly cost for the instance divided by the instance score. A lower score means better price-performance. This was changed to from the inverse in V1 to align better with conventional definitions.
Performance by Cloud Service Provider

615% performance difference between the lowest and highest performing 1 core instances.

The first view of the benchmark results looks at the range of performance by cloud service provider. The details of how the numbers were generated can be found in Appendix 1. The X-axis is the instance type, the Y axis is the relative performance against the Burststorm Compute Unit (BCU) calculated as an average over all available data points.

Figure 7: Amazon AWS
Figure 8: Google Compute Engine

Figure 9: Digital Ocean

Figure 10: Linode
Figure 11: Microsoft Azure

Figure 12: Rackspace

Figure 13: Softlayer
Amazon AWS has the largest variety of options equaled by Azure and close followed by Rackspace. Microsoft Azure now has the highest performing instance type. Taking the top spot from AWS. The interactive report will allow you to compare different suppliers, and different instance types over a larger variety of vectors if you want to dig deeper.

We noted a lot of cloud service providers have similar performance scores for different instance types. We believe this to be a function of 2 variables that play here: The UnixBench performance score does not show a lot of impact from different memory sizes and the differences in IO capability of the instance type. The latter becomes clearer when you look at Amazon AWS CPU scores vs IO scores: IO shows a more linear pattern.

This is also where we want to point out that Amazon AWS has a T-series instance type that has a “performance quota”. This means that as you use the instance over time you use up the quota and once used up, the performance goes down. This favors our testing method where we run 1 benchmark per instance 1x per day in less than 30 minutes, as opposed to continually testing 1 instance over a longer period of time.
You can see an example of how that looks like for 1 core systems in the picture below:

Figure 15: performance scores for 1 core instances

A notable part that has not changed much is that as a result of the diversity of platforms and solutions, our latest benchmark shows scores between 1.87 and 11.5 or a 615% performance difference between the lowest and highest performing 1 core instances. You can also see the difference between 2, 4, 8, 16, 32 and 36-core instances in the interactive report.
Absolute Performance

The current top three highest performing cloud services of all are:

**Microsoft Azure**
- **Instance:** m4.10xlarge
- **CPU:** E5-2676 v3 @ 2.40GHz
- **Cores:** 40
- **Location:** Ashburn, VA
- **Price:** $1815
- **Performance Price:** $33.78
- **Price BU:** 203.73
- **IO BU:** 8.91

**Amazon/AWS**
- **Instance:** m3.8xlarge
- **CPU:** E5-2698 V3 @ 2.00GHz
- **Cores:** 32
- **Location:** Ashburn, VA
- **Price:** $3210
- **Performance Price:** $22.13
- **Price BU:** 175.41
- **IO BU:** 9.49

**Amazon/AWS**
- **Instance:** m4.10xlarge
- **CPU:** E5-2676 v3 @ 2.40GHz
- **Cores:** 40
- **Location:** Ashburn, VA
- **Price:** $1815
- **Performance Price:** $33.78
- **Price BU:** 203.73
- **IO BU:** 8.91
Price performance

*Price performance for a 4-core compute cloud service can vary by 1501%*

The **Burstorm CAD application**'s product catalog contains product pricing and so we were able to connect a price to each of the instances. While the Burstorm application's product catalog contains many pricing models (hourly, month-to-month, 12, 24, 36 months etc.) in these results we used the hourly rate without discounts. The modeling part of the Burstorm application does consider the impact of other pricing models.

Figure 14 shows the performance and price-performance of all 4-core compute cloud services from the seven suppliers. Price-performance scores are calculated by dividing the price/month by the performance score. The lowest (best) score is $2.27/BCU and the highest is $34.07/BCU representing a 1501% difference. Significantly up from last time.

![Figure 16: Price Performance of 4 core instances](image)
The top three cloud services for 4 core instances by price-performance are now:

<table>
<thead>
<tr>
<th>Cloud Service</th>
<th>Instance Type</th>
<th>Model</th>
<th>Price</th>
<th>Price + Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linode</strong></td>
<td>Linode-4GB</td>
<td>Intel(R) Xeon(R) CPU E5-2660 v2 @ 2.80GHz</td>
<td>$43</td>
<td>2.27</td>
</tr>
<tr>
<td><strong>Digital Ocean</strong></td>
<td>8gb</td>
<td>Intel(R) Xeon(R) CPU E5-2660 V2 @ 2.30GHz</td>
<td>$86</td>
<td>5.28</td>
</tr>
<tr>
<td><strong>Rackspace</strong></td>
<td>general1-4</td>
<td>Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz</td>
<td>$115</td>
<td>6.81</td>
</tr>
</tbody>
</table>
The best price-performance 4-core compute cloud services is the Linode-4GB ($2.27/BCU) and is about 15x better than the price-performance from AWS's i2.xlarge ($34.07/BCU). In fact, the below graph shows that the 'Linode-4GB' is still about 2.5x better than the number 2, Digital Oceans ‘8GB’. Since the 1st version of this paper Linode upgraded their virtualization layer which preserved their price/performance lead and we included more Rackspace instances which are reflected in this updated version. The constant changes are clearly visible in the continuous benchmark application and are available to follow in the interactive report.

![Figure 17: Price performance for 4-core instance types](image)

As you can see, normalizing to price by performance can significantly change the picture.
Even within a provider the economic impact can be a key differentiator:

You can see how similar systems from a performance perspective have almost a 450% difference in price performance. The ability to see this impact helps ask questions of what is really different and relevant.

If your workload can be distributed over multiple instances looking at price performance is critical to finding the right instance for you. Prices change regularly so this is something you want to monitor over time and adjust accordingly. Since we bind the price to the data point in time the interactive report shows not only the current price-performance but also how the price-performance changes for an instance type over a period of time.
Performance over time

The same instance performance can still fluctuate by 62% over time

We have been benchmarking continually since the initial release of this report. As you can see in the below, performance of a particular compute cloud service can vary over time. The next charts show the changes in performance over time by cloud service provider. Each data point in time is the average of all locations performance results for that instance type.

Figure 19: AWS performance over time
Figure 20: Google Performance over time

Figure 21: Digital Ocean Performance over time
Figure 22: Linode Performance over time

Figure 23: Microsoft Azure Performance over time
Figure 24: Rackspace Performance over time

Figure 25: Softlayer Performance over time

Performance is most often fairly stable over time but still could vary by as much as 62% within a single instance type (see figure 24). Generally, the CPU volatility is less than the IO volatility and volatility looks worse at the higher performing instances.
Another interesting observation is that size matters in 2 ways: Performance volatility seem to be more prevalent at the larger instance types and the larger providers like AWS, Google, Rackspace and MSFT seem to provide more consistent performance over time than the smaller providers.
To show another perspective we looked at the most active vendors by price performance: The below shows Amazon AWS, Microsoft Azure, Rackspace and Google Compute Engine 4 core Price/Performance by instance, over time.

![Price/Performance of 4 core instances of AWS, MSFT, Google and Rackspace over time](image)

It seems that normalized by performance, the race to the bottom has not yet started. Meaning that the $/BCU looks consistent which indicates that if prices went down, performance seems have gone down too, or neither happened.

For those of you interested to find out the root cause for specific instances, the interactive report is continually monitoring these metrics.

You can see that performance over time matters and it shows that there are significant differences that can impact what the ideal profile for a specific workload is, based on when it is tested.
Global observations

Not all locations are created equal in availability and performance of instance types

We spread the testing over 3 locations for each provider in 3 geographies: NA (North America), APAC (Asia) and EMEA (Europe) to benchmark performance and price-performance based on locality. Note that the results are and average of all data points collected. If no data points were collected it means it had a 100% failure to start which most often means ‘not available’ but could also mean a systemic failure in our tooling. We are continually working with suppliers to diagnose any of these events.

Here are the screenshots from the interactive report:

Figure 28: Amazon AWS regional performance
Figure 29: Google regional performance

Figure 30: Digital Ocean Performance by region
Figure 31: Linode performance by region

Figure 32: Microsoft Azure performance by region
Figure 33: Rackspace performance by region

Figure 34: Softlayer performance by region
Although the performance differences between regions are typically not extreme they are noticeable. Most noticeable is that not all instances are available everywhere. Not all high performance Microsoft Azure instance types are available in APAC for example and the G4 & 5, which are new, seemed not yet available everywhere either.

However, as a general observation, consistency seems to improve. Comparing 8 Core compute cloud services of Google Compute Engine between current (V2) and last (V1) report shows:

We suspect that time improves consistency of performance capability. The performance looks a lot more consistent now than it was then.

But there is one exception to the rule: Google 32 core instances.
Today’s data for both Google and Azure shows that Google has no more failures to launch in any region where MSFT still has some issues in same areas as in June 2015:

At Microsoft Azure the A8, A9 and (newer) A10 don’t seem to be available in APAC. For Google the 32core systems seem to perform noticeably less in APAC than any other system.

The data shows there are performance differences between regions for the same instance types within a vendor and not all instances are available in every region. The interactive report continually updates as new instance become available in different locations.
Conclusions

This update (V2) to the first comprehensive and continuous price-performance benchmark has yielded some interesting observations:

- **Performance of 1-core instances can still vary by 615% between providers.**

- **MSFT now has the top performer spot with their G5 instance,** followed by AWS’s m4.10xlarge and c3.8xlarge.

- **Price performance for a 4-core compute cloud service can vary by 1501%**
  The top three price-performance winners for 4 core systems were Linode-4GB, Digital Ocean’s 8GB instance and Rackspace’s General 1-4.

- **The same instance performance can still fluctuate by 62% over time.**
  However, it seems that size matters in 2 ways: More performance volatility seem to be more prevalent at the larger instance types and the larger providers like AWS, Google, Rackspace and MSFT seem to provide more consistent performance over time than the smaller providers.

- **Not all locations are created equal in availability and performance of instance types.** Most noticeable is that not all instances are available everywhere. Not all high performance Microsoft Azure instance types are available in APAC for example and the G4 & 5, which are new, seemed not yet available everywhere either. However, as a general observation, consistency seems to improve. Comparing 8 Core compute cloud services of Google Compute Engine between current (V2) and last (V1) report shows considerably less difference.

The rate of change in instance types, pricing, performance over time and availability of services by location confirms that the traditional way of benchmarking a small set of instance types in a unique event is not sufficient anymore in today’s world of cloud computing. To see the longer term trends and understand the wide variety of results we created the [interactive report](#) for our customers.

Continuous and comprehensive benchmarking of existing and new cloud services will reveal useful information for both suppliers and consumers of compute & storage cloud services. Rice and Burststorm will continue to expand the scope of the benchmarking and work with enterprises, academics and cloud service providers to add to our collective understanding of the cloud. If you have any questions please contact [Edward Wustenhoff](#).
Appendix 1: Test details

Test details

Burstorm used the standard UnixBench score but scaled it to a more modern processor instead of the original SparcStation. The tests itself were not altered for this version of the benchmark so as to establish a widely understood and vetted baseline.

UnixBench is the original BYTE UNIX benchmark suite, updated and revised by many people over the years. The purpose of UnixBench is to provide a basic indicator of the performance of a Unix-like system; hence, multiple tests are used to test various aspects of the system's performance. These test results are then compared to the scores from a baseline system to produce an index value, which is generally easier to handle than the raw scores. The entire set of index values is then combined to make an overall index for the system.

For more information, you can review the project website here: https://github.com/kdlucas/byte-unixbench

Each run we spun up an instance with default settings (no optimizations) and the test data we collected is from a full UnixBench test with the iteration count set to 1. Each instance tested generated two entries, one with a single core and another with the maximum number of cores on the instance (up to 36 cores). If the instance only had one core, just one entry was generated.

BCU: System Specs

Burstorm uses the standard UnixBench score but scaled to a more modern processor and bus (a Raspberry Pi 2, ARM7 @900Mhz) from the original SparcStation 20-61 "George".
Appendix 2: What’s next

We normalized the values against the score of a Raspberry Pi 2, ARM7 @900Mhz and thus provide a relative score to focus on relative performance more than on absolute performance. This was done because we have brought this benchmark data into the Burstorm CAD application to optimize design decisions by performance and price-performance.

We realize that UnixBench provides a particular test of a UNIX system but it is widely accepted as a measurement for relative performance. We intent to enhance the I/O section because of the potential impact of larger CPU Cache and SSDs on the current tests. As part of RB-B we are considering adding benchmarks for Memory and Network. The first because we see that UnixBench seems only marginally impacted by additional memory while we know certain workloads clearly benefit from memory. The Network aspect is very interesting as it is the most widely shared resource and likely the most volatile. Also it is the most complex to test since by definition a network has dependencies on distance (within the VM, within the OS, within the system, within the local network and so forth and so on). We have plans in progress but welcome contributions from the community.

We are also continually adding more providers to the benchmark. The current Burstorm product catalog has already identified 1075+ compute & storage cloud services providers. Beyond those we’re working with enterprises and providers to benchmark private and dedicated (bare metal) compute & storage cloud services.

The longer term vision for the benchmark framework is to include multi instance benchmarks. Because the Burstorm CAD application is designed to define a complete architecture we see the possibility to then deploy it and run the RB-Benchmark on it to get an overall view of the relative performance of such design. This is obviously a complex goal and will take some time to evolve. In version 4.2 we already added some underlying capabilities to do so. Not in the least the ability to define test runs of multiple systems as a visual design within the application.

If you have any questions please contact Edward Wustenhoff.
Review History – V1

We’d like to thank all the reviewers below as well as those who chose to remain anonymous for their contributions to the V1 report.

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
<th>Affiliation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ravi Anadwali</td>
<td>Senior Manager</td>
<td>Splunk</td>
</tr>
<tr>
<td>Darren Bibby</td>
<td>VP, Channels &amp; Alliances</td>
<td>IDC</td>
</tr>
<tr>
<td>Mauricio Carreno</td>
<td>Senior Manager</td>
<td>Accenture Mexico</td>
</tr>
<tr>
<td>Larry Carvalho</td>
<td>Lead Analyst</td>
<td>IDC</td>
</tr>
<tr>
<td>Adrian Cockcroft</td>
<td>Technology Fellow</td>
<td>Battery Ventures</td>
</tr>
<tr>
<td>Mac Devine</td>
<td>VP, CTO</td>
<td>IBM</td>
</tr>
<tr>
<td>Angel Luis Diaz</td>
<td>IT Specialist, Infrastructure</td>
<td>IBM</td>
</tr>
<tr>
<td>Mark Egan</td>
<td>Partner</td>
<td>Stratafusion</td>
</tr>
<tr>
<td>Jim Enright</td>
<td>Director of Performance</td>
<td>Avnet</td>
</tr>
<tr>
<td>Tim Fitzgerald</td>
<td>VP Cloud</td>
<td>IBM</td>
</tr>
<tr>
<td>Sandeep Gopisetty</td>
<td>Distinguished Engineer</td>
<td>Dell</td>
</tr>
<tr>
<td>Dave Hansen</td>
<td>VP and General Manager</td>
<td>IBM</td>
</tr>
<tr>
<td>Andrew Hately</td>
<td>CTO</td>
<td>VMware</td>
</tr>
<tr>
<td>Bill Heil</td>
<td>SVP, Chief Bottle Washer</td>
<td>Fidelity Investments</td>
</tr>
<tr>
<td>Kristopher Johnston</td>
<td>Director IT</td>
<td>Ingram Micro</td>
</tr>
<tr>
<td>Sam Kamal</td>
<td>Global Technology Executive</td>
<td>IBM</td>
</tr>
<tr>
<td>Sunil Kamath</td>
<td>Dir. Performance Engineering</td>
<td>Zype</td>
</tr>
<tr>
<td>Ed Laczynski</td>
<td>Co-Founder and CEO</td>
<td>SAIC</td>
</tr>
<tr>
<td>Cary Landis</td>
<td>Solutions Architect</td>
<td>Microsoft</td>
</tr>
<tr>
<td>Charles Levine</td>
<td>Principal Program Manager</td>
<td>Singapore Mngt University</td>
</tr>
<tr>
<td>Dan Ma</td>
<td>Assistant Professor</td>
<td>Forrester Research</td>
</tr>
<tr>
<td>William Martroelli</td>
<td>Principal Analyst</td>
<td>Red Hat</td>
</tr>
<tr>
<td>Michael McCain</td>
<td>Enterprise Architect</td>
<td>Estee Lauder</td>
</tr>
<tr>
<td>Justin Mennen</td>
<td>VP Enterprise Architecture</td>
<td>Save the Children</td>
</tr>
<tr>
<td>Ken Murdoch</td>
<td>VP IT &amp; Bldg Operations</td>
<td>Facebook</td>
</tr>
<tr>
<td>Thao Nguyen</td>
<td>Engineer - OCE</td>
<td>Purdue University</td>
</tr>
<tr>
<td>Sanjay Rao</td>
<td>Associate Professor</td>
<td>Kaiser Permanente</td>
</tr>
<tr>
<td>Farhad Shafa</td>
<td>Solutions Architect</td>
<td>Originate</td>
</tr>
<tr>
<td>Lloyd Taylor</td>
<td>CIO</td>
<td>University of Oxford</td>
</tr>
<tr>
<td>David Wallom</td>
<td>Associate Professor</td>
<td>Constellation Research</td>
</tr>
<tr>
<td>Ray Wang</td>
<td>Principal Analyst</td>
<td></td>
</tr>
</tbody>
</table>