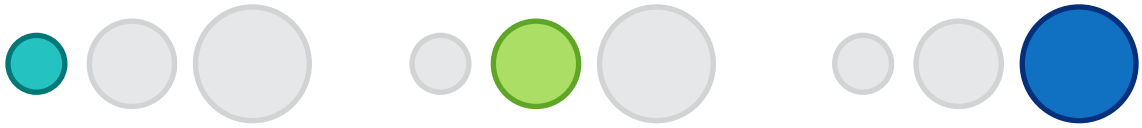
Small instances: Process data at up to 1.57x the rate

Medium instances: Process data at up to 1.42x the rate

Large instances: Process data at up to 1.72x the rate



Achieve higher data throughput for your Apache Spark machine learning workloads with M5n instances for Amazon Web Services featuring 2nd Generation Intel Xeon Scalable Processors – Cascade Lake

M5n series instances that featured 2nd Generation Intel Xeon Scalable processors handled more data per second compared to older M4 instances that featured Broadwell processors

If your company has decided to run its large-scale machine learning workloads on Amazon Web Services, you have a big decision to make: which instances will you purchase to power your work?

At Principled Technologies, we used two machine learning workloads from the HiBench benchmarking suite to test two series of small, medium, and large Amazon Elastic Cloud Compute (Amazon EC2) instances running Apache Spark. We compared newer M5n series instances that featured 2nd Generation Intel Xeon Scalable processors—also known as Cascade Lake—to older M4 series instances that featured Intel Xeon E5 v4 processors—also known as Broadwell. We found that at all three instance sizes, the newer M5n instances processed a higher volume of data compared to the older instances.

These advantages could help your business to organize its data and complete big data analysis work in less time, enabling you to reach mission-critical conclusions faster. This could in turn lead to key decisions that could give your business the edge over competitors.

How we tested

We tested two series of general-purpose Amazon EC2 instances:

- Newer M5n series instances featuring Cascade Lake processors
- Older M4 series instances featuring Broadwell processors

We compared these instances across three sizes: small instances with 8 vCPUs, medium instances with 16 vCPUs, and large instances with 64 vCPUs. For each environment, we created a five-node cluster of identical instances.

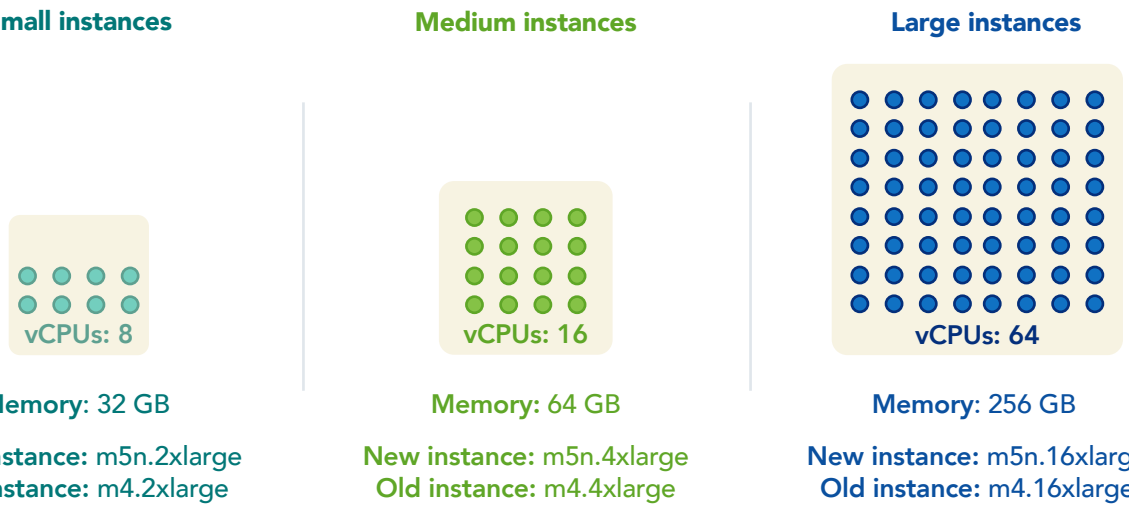


Figure 1: Specifications for the Amazon EC2 instances we used. We tested each instance in the us-east-1f region. Source: Principled Technologies.

Testing with the HiBench benchmark suite

We assessed the Amazon EC2 instances with two machine leaning implementations from the HiBench benchmark suite:

- The training process for Naive Bayesian classification
- The k-means clustering algorithm

Large-scale machine learning involves processing and analyzing unstructured data from many sources. Bayesian classification assesses a system's ability to mine data to make probabilistic predictions, while the k-means algorithm uses data-point clustering to provide business insights, stressing the compute capabilities of a system. With each workload representing different types of data manipulation, we can assess multiple aspects of the instance cluster's machine learning performance.

Our results

Small instances

Smaller organizations that don't require a great deal of computing resources can still benefit from purchasing newer instances for their machine learning work. Figure 2 shows that the m5n.2xlarge instance cluster that featured Cascade Lake processors cluster processed a higher volume of data on average than the older m4.2xlarge instance cluster that featured Broadwell processors. This was true for both the Bayesian classification and k-means clustering workloads.

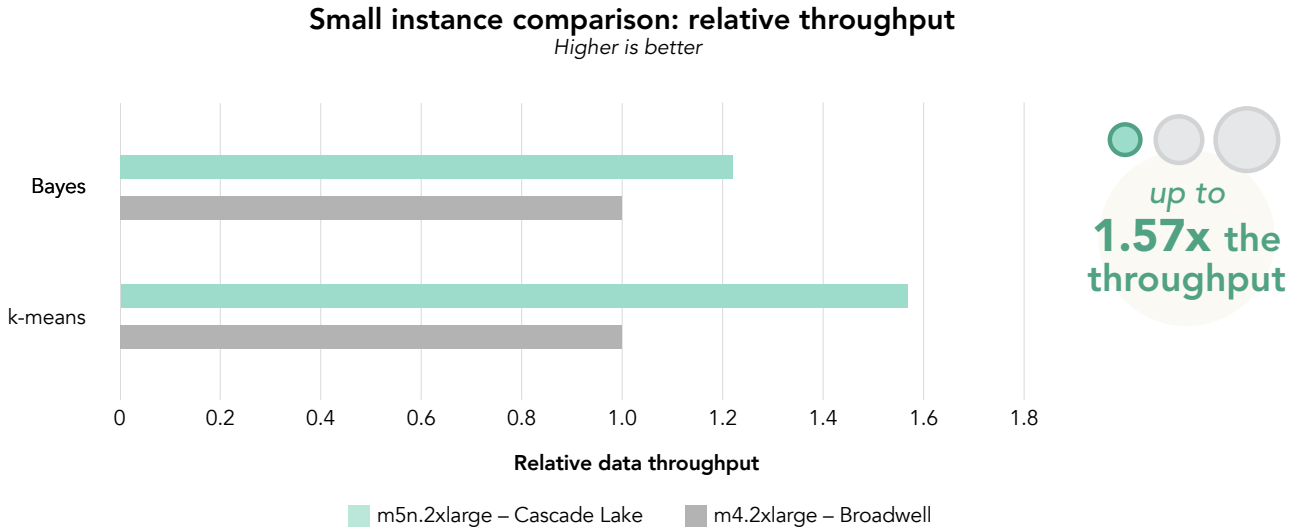


Figure 2: Normalized comparison of the average throughput each small instance cluster achieved in the Naive Bayesian classification and k-means clustering workloads. Higher throughput is better. Source: Principled Technologies.

Medium instances

Businesses with more data to analyze may find that newer medium-sized M5n instances suit their needs. Figure 3 shows that, on average, the Cascade Lake processor-powered m5n.4xlarge instance cluster processed 1.22 times the volume of data on the Bayesian classification benchmark compared to the older m4.4xlarge instance cluster that featured Broadwell processors, and 1.42 times the data throughput on the k-means clustering workload.

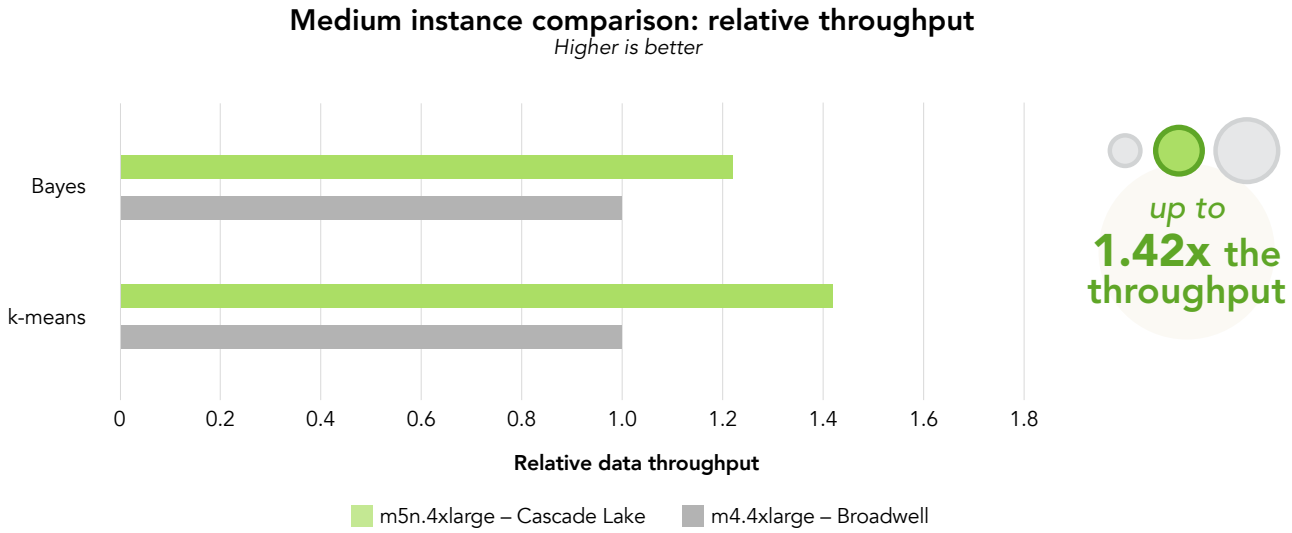


Figure 3: Normalized comparison of average data throughput each medium instance cluster achieved in the Naive Bayesian classification and k-means clustering workloads. Higher throughput is better. Source: Principled Technologies.

Large instances

Large businesses with vast stores of critical data to analyze need all the processing power they can afford. Figure 4 shows that the m5n.16xlarge instance cluster, which featured Cascade Lake processors, handled a larger volume of data than the m4.16xlarge instance cluster that featured Broadwell processors. On average, the cluster of newer instances processed 1.56 times the data per second on the Bayesian classification benchmark, and 1.72 times the data per second on the k-means clustering workload.

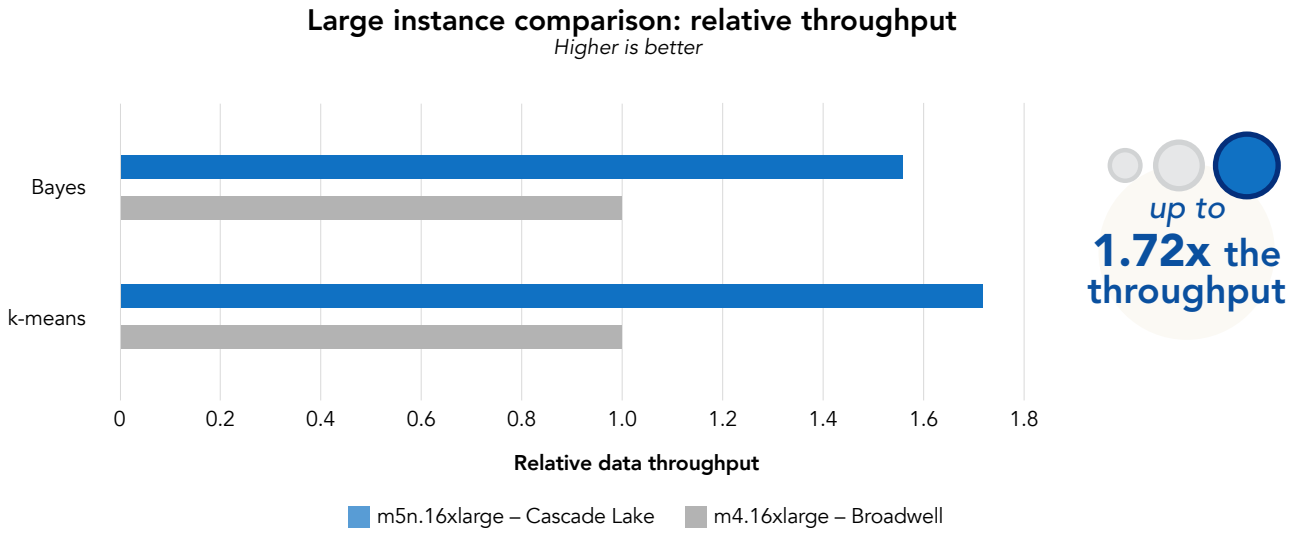


Figure 4: Normalized comparison of the average data throughput each large instance cluster achieved in the Naive Bayesian classification and k-means clustering workloads. Higher throughput is better. Source: Principled Technologies.

Better performance, better value for your investment

In our tests, the newer M5n series instance clusters handled a range of 1.22x to 1.72x higher volume of data compared to the older M4 series instance clusters. Because M5n instances cost just 1.19x as much as the older instances, they may be a better value for machine learning workloads such as the ones we tested.[1]

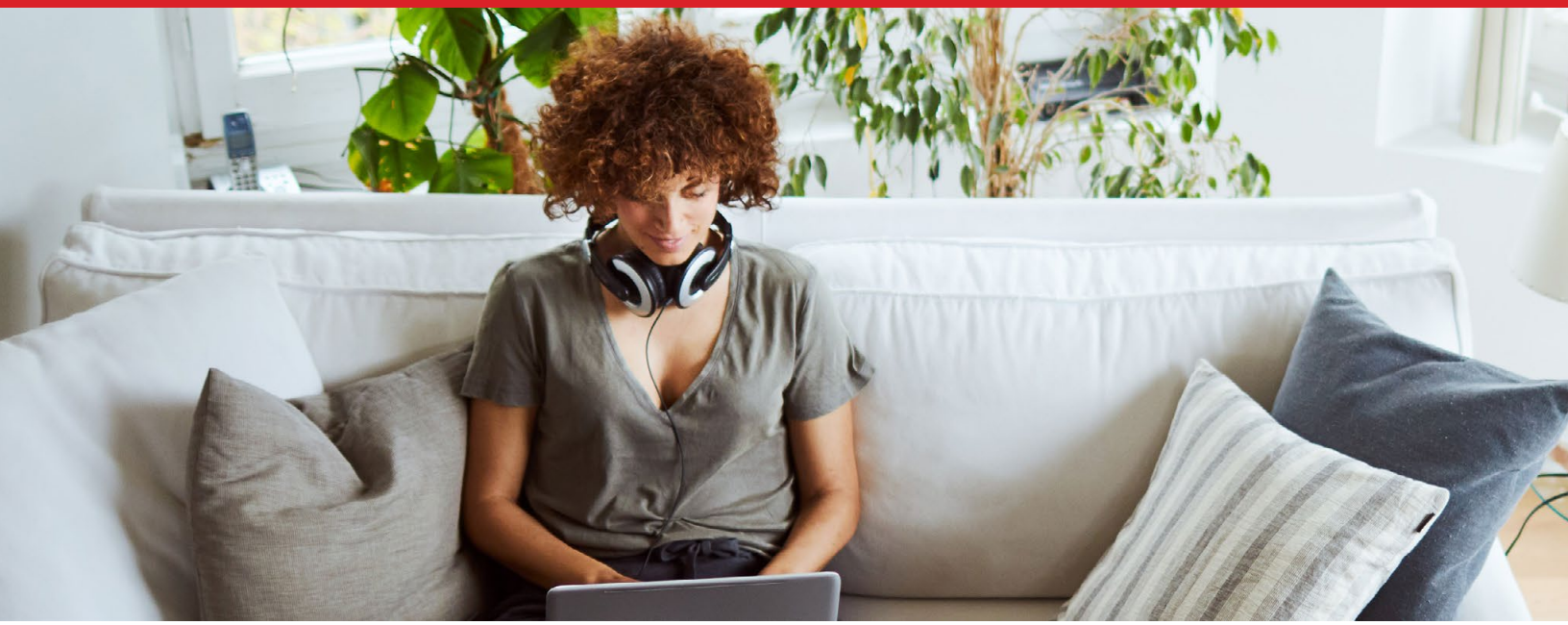
M5n series instances for Amazon EC2

In 2019, Amazon introduced M5n instances to their EC2 offerings, extending 100Gbps networking and adding improved packet-processing performance capabilities to its general-purpose M family.[2] In addition to adding 2nd Generation Intel Xeon Scalable processors to the M-series options, M5n instances support the following:

- Higher CPU core frequency (3.1 GHz vs 2.3 GHz)
- Intel DL Boost Vector Neural Network Instructions, which Intel claims can "deliver a significant performance improvement" for deep learning applications[3]
- 25 Gbps of peak bandwidth for small and medium instances, 75Gbps for large instances
- AWS Nitro System, which provides better security and performance compared to older AWS hypervisors[4]

Performance that scales at higher vCPU counts

In both the Bayesian classification and k-means clustering workloads, the performance margin between the M5n and M4 series instances increased significantly between the medium (16vCPU) and large (64vCPU) instance cluster tests.



Conclusion

Your organization deserves powerful instances to run its cloud-based Apache Spark machine learning work. The time you save on big data analysis can go toward informed planning and decision-making that can affect the outcome of critical business goals.

In our Naive Bayesian classification and k-means clustering tests, newer M5n series instances for Amazon EC2 that featured Cascade Lake processors handled more data per second on average compared to older M4 series instances that featured Broadwell processors. The performance advantages we saw during testing also mean that organizations that invest in the newer instances would be getting better value for their money.

---

1 "Amazon EC2 On-Demand Pricing," accessed October 28, 2020, https://aws.amazon.com/ec2/pricing/on-demand/

2 Julien Simon, "New M5n and R5n EC2 Instances, with up to 100Gbps networking," accessed November 3, 2020, https://aws.amazon.com/blogs/aws/new-m5n-and-r5n-instances-with-up-to-100-gbps-networking/.

3 "Introduction to Intel Deep Learning Boost on Second Generation Intel Xeon Scalable Processors," accessed November 3, 2020, https://software.intel.com/content/www/us/en/develop/articles/introduction-to-intel-deep-learning-boost-on-second-generation-intel-xeon-scalable.html.

4 "AWS Nitro System," accessed November 16, 2020, https://aws.amazon.com/ec2/nitro/.

Read the science behind this report at http://facts.pt/AFf9GVO ▶

Principled Technologies®

Facts matter.®

Principled Technologies is a registered trademark of Principled Technologies, Inc.
All other product names are the trademarks of their respective owners.
For additional information, review the science behind this report.

This project was commissioned by Intel.