

# Analyze More Data with up to 1.72x the Throughput for Apache Spark™ Workloads with Amazon™ EC2 M5n Instances Featuring 2<sup>nd</sup> Gen Intel® Xeon® Scalable Processors

## Amazon Web Services M5n Series Instances Feature Intel Cascade Lake Scalable Processors



Apache Spark



Analyze more data with 1.57x the throughput on small instances



1.42x the throughput on medium instances



1.72x the throughput on large instances

### Boost Throughput for Machine Learning with Amazon EC2 M5n Series Instances 2<sup>nd</sup> Gen Intel Xeon Scalable Processors

As Big Data continues to grow, organizations must find ways to sort through and harness the lessons gleaned from that data to remain agile in the marketplace. Running data analysis in the cloud offloads on-prem administration hassles, but it can make it difficult to discern the impact that the choice of instance can have on the performance of complex data analysis workloads. For Apache Spark workloads on Amazon EC2, selecting M5n instances enabled by 2<sup>nd</sup> Gen Intel Xeon Scalable processors can provide more throughput to sort through more data at a time to give you insights faster.

In tests of two machine learning implementations comparing Amazon EC2 instances, newer M5n series instances enabled by 2<sup>nd</sup> Gen Intel Xeon Scalable processors outperformed older M4 series instances with Intel Xeon E5 v4 processors, delivering up to 1.72x the data throughput for Apache Spark workloads.

Whether your machine learning workloads require, small, medium, or large instances sizes, selecting M5n series instances featuring 2<sup>nd</sup> Gen Intel Xeon Scalable processors over older M4 instances can analyze more data and deliver actionable insights faster.

### Improve Time to Insight on Small Instances

All the data an organization collects is only worthwhile if they can quickly make sense of it. For example, customer preference predictions and similar inferences must work in real time to have a business impact—and this requires updated technology that can deliver results faster.

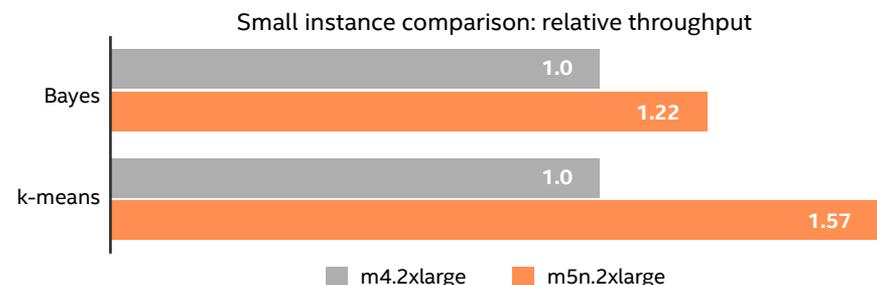


Figure 1. Relative throughput comparison on small instances (8 vCPU/32GB RAM) for Naïve Bayesian classification and k-means clustering workloads from the HiBench benchmark suite.

Tests comparing small instances with eight vCPUs show choosing Amazon EC2 M5n instances featuring 2<sup>nd</sup> Gen Intel® Xeon® Scalable processors offers up to 1.57x the throughput for Apache Spark machine learning workloads of M4 series instances with Intel Xeon E5 v4 processors.

### Improve Time to Insight on Medium Instances

As with small instances, tests comparing medium instances with 16 vCPUs showed that Amazon EC2 M5n instances featuring 2<sup>nd</sup> Gen Intel® Xeon® Scalable processors improved both machine learning implementations on Apache Spark—in this case, delivering up to 1.42x the throughput of older M4 instances.

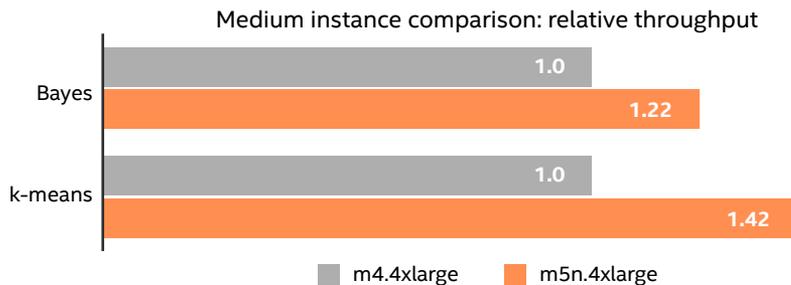


Figure 2. Relative throughput comparison on medium instances (16 vCPU/64GB RAM) for Naïve Bayesian classification and k-means clustering workloads from the HiBench benchmark suite.

### Improve Time to Insight on Large Instances

Testing shows that large instance sizes (with 64 vCPU) had the most dramatic increase in machine learning performance, offering up to 1.72x the throughput of M4 series instances for a k-means clustering workload.

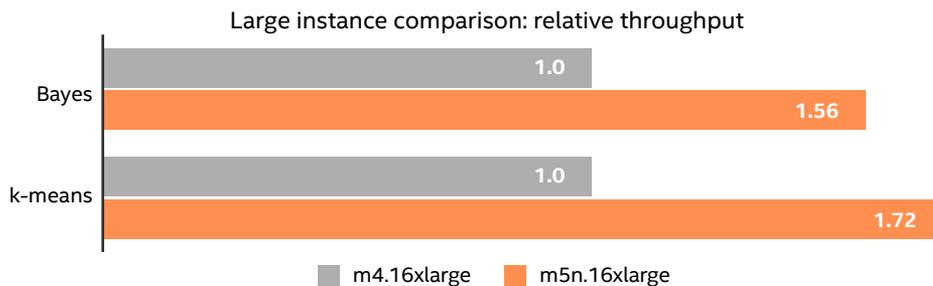


Figure 3. Relative throughput comparison on large instances (64 vCPU/256GB RAM) for Naïve Bayesian classification and k-means clustering workloads from the HiBench benchmark suite.

This means that organizations looking to quickly gain actionable insights from data can benefit from selecting upgraded Amazon EC2 M5n instances enabled by 2<sup>nd</sup> Gen Intel Xeon Scalable processors, no matter the size of the instances they require.

### Learn More

To begin your Apache Spark deployments on Amazon EC2 M5n series instances with 2<sup>nd</sup> Gen Intel Xeon Scalable processors, visit <http://intel.com/AWS>.

For more test details, visit <http://facts.pt/3Kjn66x>.



Performance varies by use, configuration and other factors. Learn more at <https://intel.com/benchmarks>.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy. Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

Printed in USA 0221/JO/PT/PDF US001

