

Analyze up to 1.55x the Data per Second for Apache Spark™ Workloads with Microsoft® Azure® Dds_v4 VMs Featuring 2nd Gen Intel® Xeon® Scalable Processors

Microsoft Azure Dds_v4 Series VMs Feature Intel Cascade Lake Scalable Processors

Apache Spark

Analyze more data per second with 1.23x the throughput
on small VMs

1.46x the throughput
on medium VMs

1.55x the throughput
on large VMs

Speed Machine Learning Workloads with Microsoft Azure Dds_v4 Series VMs Featuring 2nd Gen Intel Xeon Scalable Processors

Making sense of the massive amounts of data your organization collects is a big task—one that requires updated technology to get the job done quickly. The Microsoft Azure cloud VMs you select to host Apache Spark clusters dictate how fast you can get actionable information from your data and turn it into business strategy. For demanding Apache Spark machine learning workloads on Microsoft Azure, selecting Dds_v4 VMs enabled by 2nd Gen Intel Xeon Scalable processors can allow you to analyze more data per second to boost the agility of your business.

In tests of two machine learning implementations comparing Microsoft Azure VMs, newer Dds_v4 series VMs enabled by 2nd Gen Intel Xeon Scalable processors outperformed older Ds_v3 series VMs with Intel Xeon E5 v4 processors, analyzing up to 1.55x the data per second for Apache Spark machine learning workloads.

Across, small, medium, and large VMs sizes, selecting Dds_v4 series VMs featuring 2nd Gen Intel Xeon Scalable processors over older Ds_v3 VMs can allow you to sort more data, faster, and make quick business decisions rooted in the truth of that data.

Small Businesses Get Insights Sooner with Small VMs

Just because an organization is small doesn't mean its machine learning demands are. For large-scale machine learning needs on small VMs, choosing updated technology can ensure cloud VMs meet current demands and provide room to grow.

Tests comparing small VMs with eight vCPUs show choosing Microsoft Azure Dds_v4 VMs featuring 2nd Gen Intel Xeon Scalable processors can boost Apache Spark machine learning workloads by up to 1.23x the data per second workloads of Ds_v3 series VMs with Intel Xeon E5 v4 processors.

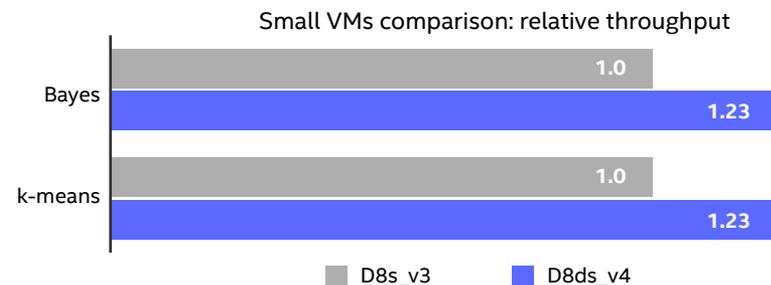


Figure 1. Relative throughput comparison on small VMs (8 vCPU/32GB RAM) for Naïve Bayesian classification and k-means clustering workloads from the HiBench benchmark suite.

Mid-Sized Businesses Get Insights Sooner with Medium VMs

As in testing with small VM sizes, tests comparing medium VMs with 16 vCPUs showed that Microsoft Azure Dds_v4 VMs featuring 2nd Gen Intel® Xeon® Scalable processors improved both Naïve Bayesian and k-means clustering machine learning implementations on Apache Spark—in this case, delivering up to 1.46x the throughput of older Ds_v3 VMs.

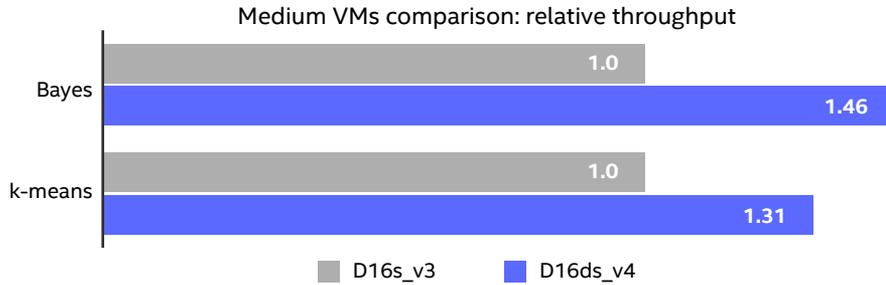


Figure 2. Relative throughput comparison on medium VMs (16 vCPU/64GB RAM) for Naïve Bayesian classification and k-means clustering workloads from the HiBench benchmark suite.

Enterprises Get Insights Sooner with Large VMs

Testing shows that the biggest throughput improvement for Apache Spark machine learning performance can with larger instance sizes (with 64 vCPUs), offering up to 1.55x the throughput of Ds_v3 series VMs for a Naïve Bayesian classification test.

Compared to the older Ds_v3 series, Microsoft Azure Dds_v4 VMs enabled by 2nd Gen Intel Xeon Scalable processors offer dramatic performance improvements, offer 50 percent larger default disk drives, and provide high IOPs on default disk drives, no matter the size of the VMs they require. This enables Azure Dds_v4 VMs to improve machine learning over the Ds_v3 series at multiple VM sizes.

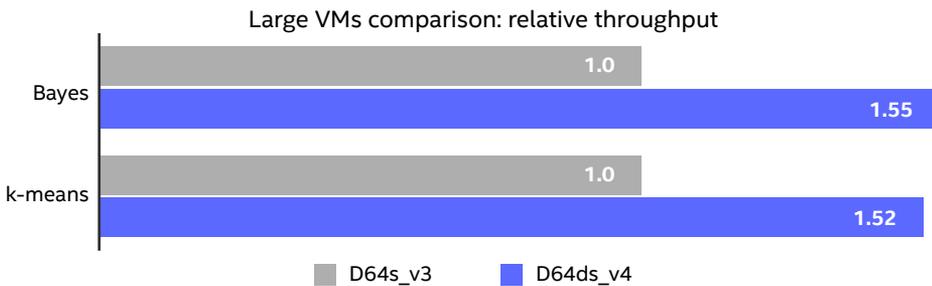
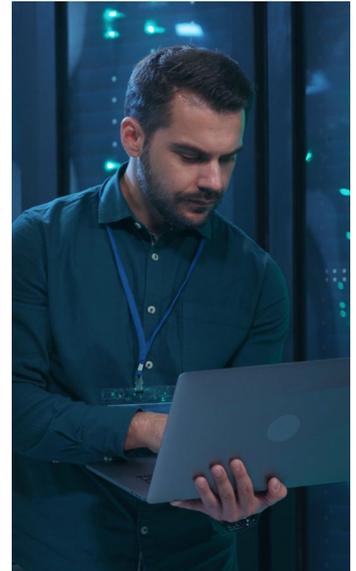


Figure 3. Relative throughput comparison on large VMs (64 vCPU/256GB RAM) for Naïve Bayesian classification and k-means clustering workloads from the HiBench benchmark suite.

Learn More

To begin your Apache Spark machine learning workloads on Microsoft Azure Dds_v4 series VMs with 2nd Gen Intel Xeon Scalable processors, visit <http://intel.com/Azure>.

For more test details, visit <http://facts.pt/pg16MAO>.



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

Please Recycle