



AWS M5 Instances Featuring 2nd Gen Intel[®] Xeon[®] Scalable Processors Speed Up Amazon EMR Big Data Analysis by up to 40%

Compared to M5a Instances with AMD EPYC Processors Across Several Instance Sizes

The Amazon EMR platform (formerly Amazon Elastic MapReduce) allows organizations to simplify running big data frameworks on AWS instances. Choosing an instance type with more powerful processors can speed up data analysis and help your bottom line. Using the TPC-DS 2.4 benchmark, we measured the EMR performance of several Amazon Web Services (AWS) EC2 cloud instances. We found that both medium-sized and larger M5 instances enabled by 2nd Gen Intel Xeon Scalable processors sped up EMR data analysis compared to same-size M5a instances with AMD EPYC processors.

Based on these test results across instance sizes, organizations seeking to speed EMR workloads (which include Apache Spark 3.1.1 and Hadoop 3.2.1) for quicker data analysis could gain insights faster by selecting AWS M5 instances featuring 2nd Gen Intel Xeon Scalable processors.

Improve Amazon EMR Performance by up to 31% on Medium-Sized Instances

For instances with 16 vCPUs, the m5.4xlarge instance enabled by 2nd Gen Intel Xeon Scalable processors improved Amazon EMR performance by up to 31% compared to the m5a.4xlarge instance with AMD EPYC processors (see Figure 1). Similarly, at 8 vCPUs, the m5.2xlarge instance improved big data analysis over the m5a.2xlarge instance by up to 19%.

Relative Completion Speed for Medium-Sized Instances

Relative Speed | Higher is better

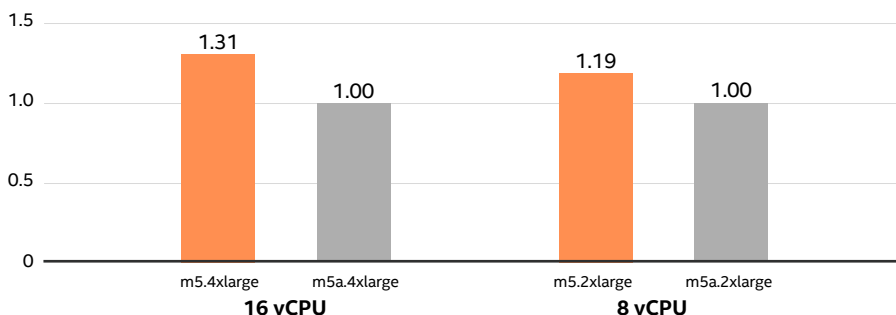




Figure 1. Relative speed of Amazon EMR analytics workloads at 16 and 8 vCPUs achieved by M5 instances with 2nd Gen Intel Xeon Scalable processors vs. M5a instances with AMD EPYC processors.

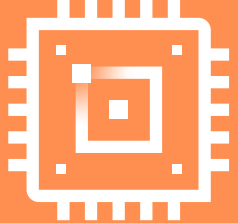


Spark



Speed Up EMR Performance by up to 31% by Selecting AWS m5.4xlarge Instances Featuring 2nd Gen Intel Xeon Scalable Processors

vs. m5a.4xlarge Instances with AMD EPYC Processors



Speed Up EMR Performance by up to 40% by Selecting AWS m5.8xlarge Instances Featuring 2nd Gen Intel Xeon Scalable Processors

vs. m5a.8xlarge Instances with AMD EPYC Processors

Improve Amazon EMR Performance by up to 40% on Larger Instances

As Figure 2 shows, comparing instances with 48 vCPUs, the m5.12xlarge instance enabled by 2nd Gen Intel® Xeon® Scalable processors sped up Amazon EMR performance by up to 28% compared to the m5a.12xlarge instance based on AMD EPYC processors. At 32 vCPUs, the m5.8xlarge instance sped up analysis over the m5a.8xlarge instance by 40%.

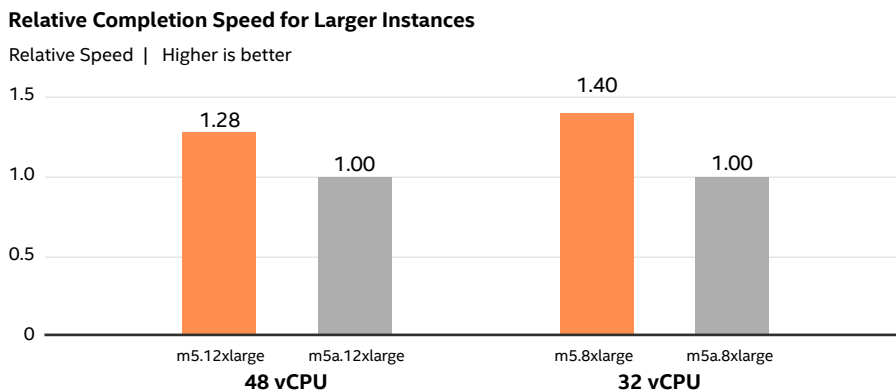


Figure 2. Relative speed of Amazon EMR analytics workloads at 48 and 32 vCPUs achieved by M5 instances with 2nd Gen Intel Xeon Scalable processors vs. M5a instances with AMD EPYC processors.

Conclusion

Across the four instance sizes we tested, AWS M5 instances featuring 2nd Gen Intel Xeon Scalable processors sped up Amazon EMR performance compared to same-sized AMD EPYC processor-based AWS M5a instances. These results show that organizations hosting big data platforms on AWS can speed up data analysis and get insights faster by selecting AWS M5 instances with 2nd Gen Intel Xeon Scalable processors.

Learn More

To begin running your Amazon EMR analytics workloads on M5 instances with 2nd Gen Intel Xeon Scalable processors, visit <https://aws.amazon.com/ec2/instance-types/M5/>.

Test by Intel in Jan. 2022. Tests on AWS us-east-1 with Linux 4.14.225-169.362.amzn2.x86_64 #1 SMP, EMR 6.3.0, Apache Spark 3.1.1, and Hadoop 3.2.1. All AMD VMs with AMD EPYC 7571. Instance details: m5.12xlarge: 5x nodes, Intel Xeon 8175M, 48 vCPUs, 192GB RAM, EBS 512GB, 10Gbps NW BW, 9,500 Mbps Storage BW; m5.8xlarge: 5x nodes, Intel Xeon 8259 CL, 32 vCPUs, 128GB RAM, EBS 512GB, 10 Gbps NW BW, 6,800 Mbps Storage BW; m5.4xlarge: 5x nodes, Intel Xeon 8259CL, 16 vCPUs, 64GB RAM, EBS 256GB, 10 Gbps NW BW, 4,750 Mbps Storage BW; m5.2xlarge: 10x nodes, Intel Xeon 8259CL, 8 vCPUs, 32GB RAM, EBS 128GB, Up to 10Gbps NW BW, up to 4,750 Mbps Storage BW; m5a.12xlarge: 5x nodes, 48 vCPUs, 192GB RAM, EBS 512GB, 10 Gbps NW BW, 6,780 Mbps Storage BW; m5a.8xlarge: 5x nodes, 32 vCPU, 128GB RAM, EBS 512GB, Up to 10Gbps NW BW, 4,750 Mbps Storage BW; m5a.4xlarge: 5x nodes, 16 vCPUs, 64GB RAM, EBS 256GB, Up to 10Gbps NW BW, 2,880 Mbps Storage BW; m5a.2xlarge: 10x nodes, 8 vCPUs, 32GB RAM, EBS 128GB, Up to 10Gbps NW BW, Up to 2,880 Mbps Storage BW.

