AMD DATACENTER SOLUTIONS

AMD EPYC™ and HYCOM

Powerful Options for Sophisticated Weather Modeling

Solution Brief
April, 2019

Exceptional Memory Bandwidth

AMD EPYC™ server processors deliver 8 channels of memory with support for up to 2TB of memory per processor.

Standards Based

AMD is committed to industry standards, offering you a choice in x86 processors with design innovations that target the evolving needs of modern datacenters.

No Compromise Product Line

Compute requirements are increasing, datacenter space is not. AMD EPYC server processors offer up to 32 cores and a consistent feature set across all processor models.

Power HPC Workloads

Tackle HPC workloads with leading performance and expandability. Accelerate your workloads with up to 33% more PCI Express® Gen 3 lanes for high performance devices, including Mellanox InfiniBand adapters.

Optimize Productivity

Increase productivity with tools, resources, and communities to help you "code faster, faster code." Boost application performance with Software Optimization Guides and Performance Tuning Guidelines.

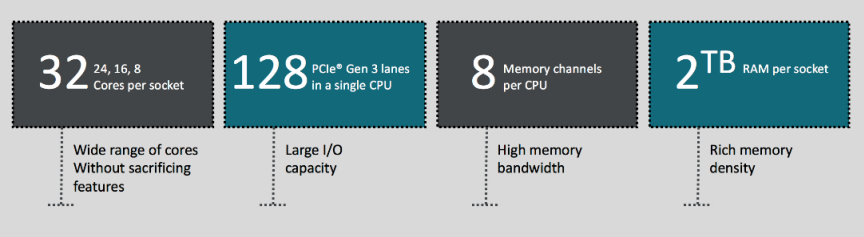
Security Features

Help safeguard your software and data with the industry's first x86 processor with an embedded security processor.

AMD EPYC: A Philosophy of Choice for High Performance Computing

Designed from the ground up for a new generation of solutions, AMD EPYC implements a philosophy of choice without restriction. Choose the number of cores and sockets that meet your needs without sacrificing key features like memory and I/O.

Each EPYC processor can have from 8 to 32 cores with access to an exceptional amount of I/O and memory regardless of the number of cores in use, including 128 PCIe® lanes, and access to up to 2 TB of high speed memory per socket.



The AMD EPYC processor's innovative architecture translates to tremendous performance at a low cost. More importantly, the performance you're paying for is appropriate to the performance you need.

Sophisticated Weather Modeling using AMD EPYC and HYCOM

In this paper, we describe the performance characteristics of AMD EPYC processors with different frequencies and core counts when running benchmarks from the HYCOM weather application.

HYCOM, the HYbrid Coordinate Ocean Model, combines specialized coordinate systems for the deep ocean, the mixed layer nearer the surface and the terrain-following, shallow coastal regions. This combination improves the relevancy of the forecast over previous models, but increases the level of complexity of the fluid dynamic equations involved that govern the interaction of wind, sea, and land over large volumes of the Earth.

HYCOM Benchmarks

The HYCOM application provides various performance metrics in the output it generates when running these benchmarks. Of the output generated, we have focused on the wall clock time taken for the overall run.

The benchmark cases chosen for this paper were two global weather simulations. The benchmark cases are called GLBT0.72 and GLBT0.08, where "GLB" stands for Global HYCOM Benchmark:

- These first benchmark, GLBT0.72, is smaller and easily runs on a single compute node. It was chosen to help tune our installation process, while still providing useful data regarding HYCOM's behavior on EPYC processors.
- The second benchmark, GLBT0.08, is a more challenging workload and is equivalent to a true production run.

Figure 1 shows the total heat flux for GLBT0.08. The specific visualization shown here is for illustration purposes only, to help the reader understand some of the output produced by these benchmark runs.

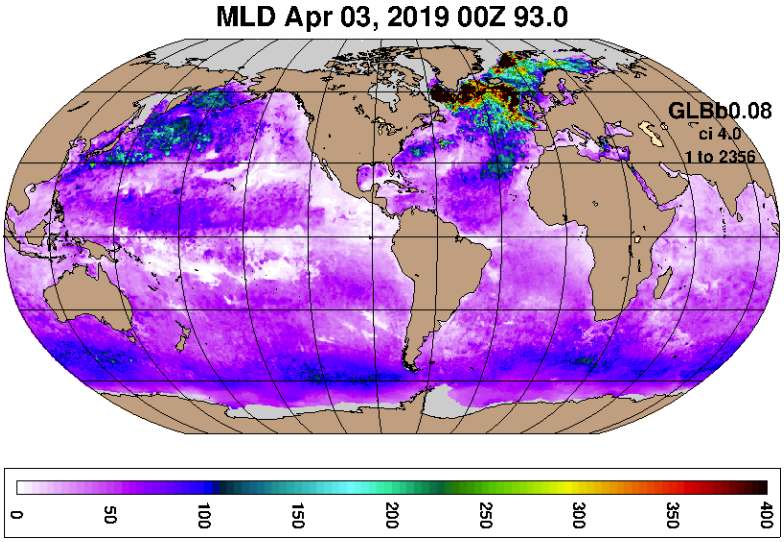


Figure 1: Representative output for HYCOM's GLBT0.08 benchmark showing a global "nowcast"[1]

The smaller of the test cases, GLBT0.72, is a reduced version of the larger test case, GLBT0.08. The numbers after the periods in the test case designations refer to the number of degrees of latitude for the ocean and ice grid. These correspond to extents in kilometers of the underlying computational grid, e.g., 0.72º corresponds to 75 km. See Table 1.

Grids	Small Test Case: GLBT0.72	Large Test Case: GLBT0.08
Ocean & ice grid	0.72º	0.08º
Atmosphere & land	1.9º x 2.5º	0.47º x 0.63º

Table 1: Global HYCOM benchmark details

Test Hardware & Software Configuration

HYCOM testing was performed on three separate 8-node clusters. Each cluster is composed of dual socket nodes:

- cluster 1 is composed of 2 x EPYC™ 7451 processors with 24 cores/socket, or 48 cores/node;
- cluster 2 is composed of 2 x EPYC™ 7351 processors with 16 cores/socket, or 32 cores/node;
- cluster 3 is composed of 2 x EPYC™ 7371 processors with 16 cores/socket, or 32 cores/node,

HYCOM testing in this paper was performed using the Message Passing Interface (MPI) with increasing MPI rank counts. MPI is a commonly used form of distributed computing, in which a program spawns MPI ranks, i.e. full programs with private memory spaces.

Compute Nodes	
CPUs: Cluster 1	2 x EPYC 7451 processors, 24 cores / socket, 48 cores / node 2.3 GHz base / 2.9 GHz all core boost
CPUs: Cluster 2	2 x EPYC 7351 processors, 16 cores / socket, 32 cores / node 2.4 GHz base / 2.9 GHz all core boost
CPUs: Cluster 3	2 x EPYC 7371 processors, 16 cores / socket, 32 cores / node 3.1 GHz base / 3.6 GHz all core boost
Cache	512 KB L1D$/core; 8 MiB L2$/ core complex
Memory	256GB Dual-Rank DDR4-2666; 8 channels per processor.
NIC	Mellanox ConnectX-5 EDR 100Gb Infiniband x16 PCIe®
Storage: OS	1 x 256 GB NVMe
Storage: Data	1 x 1 TB NVMe
Software	
OS	RHEL 7.5 (3.10.0-862.el7.x86_64)
Mellanox OFED Driver	MLNX_OFED_LINUX-4.3-3.0.2.1 (OFED-4.3-3.0.2)
MPI Version	OpenMPI 4.0.0
Application	HYCOM 2.2
Network	
Switch	Mellanox EDR 100Gb/s Managed Switch (MSB7800-ES2F)
Configuration Options	
BIOS Setting	SMT=OFF, Boost=ON, Determinism Slider = Power
OS Settings	Transparent Huge Pages=ON (Default), Swappiness=0, Governor=Performance

Table 2: Hardware & Software Test Configuration

Benchmark Results: GLBT0.72

For the GLBT0.72 benchmark we used the AMD EPYC 7451 processor, which has 24 cores, a base frequency of 2.3 GHz, and all core boost frequency of 2.9 GHz (see Table 2).

Note that HYCOM has prescribed MPI rank counts for its test cases. In practical terms, this means that one cannot choose arbitrary MPI rank counts for HYCOM MPI runs. Available rank decompositions for the GLBT0.72 test case are 16, 32, 46 (not 48 as one might expect) and 64 ranks.

MPI ranks are mapped to cores. Cluster 1's nodes are EPYC 7451s each with a total of 48 cores. As such, this test case's highest MPI rank count of 64 cannot fit on a single Cluster 1 node. Therefore, a transition from intra-node to inter-node communication must occur at the 64 MPI rank mark. **Error! Reference source not found.** shows five run averages, with the speedup computed relative to the lowest core count run (16-cores). The transition does not visibly affect the scaling curve.

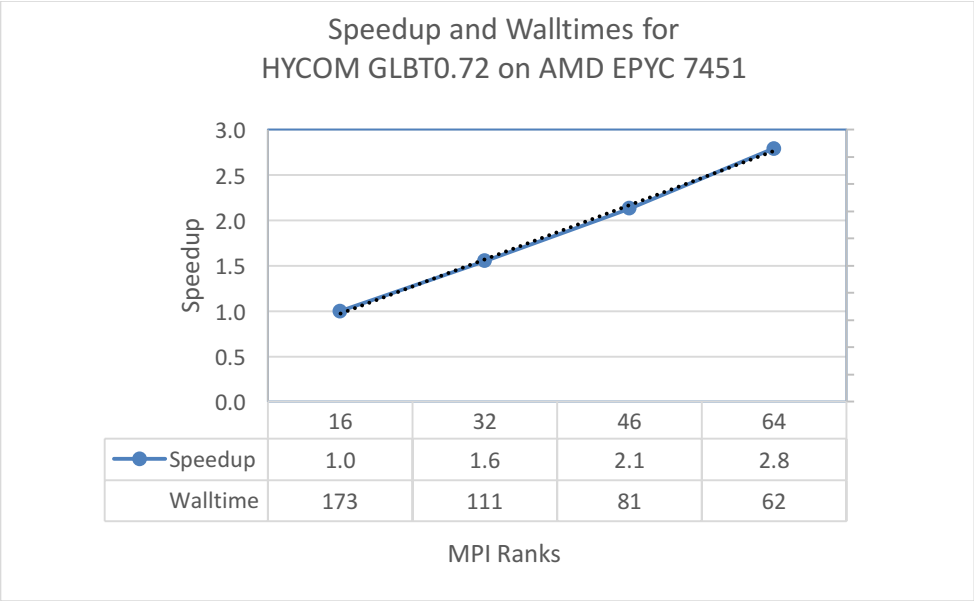


Figure 2: Lower resolution test case GLBT0.72 using EPYC 7451 cluster.

Benchmark Results: GLBT0.08

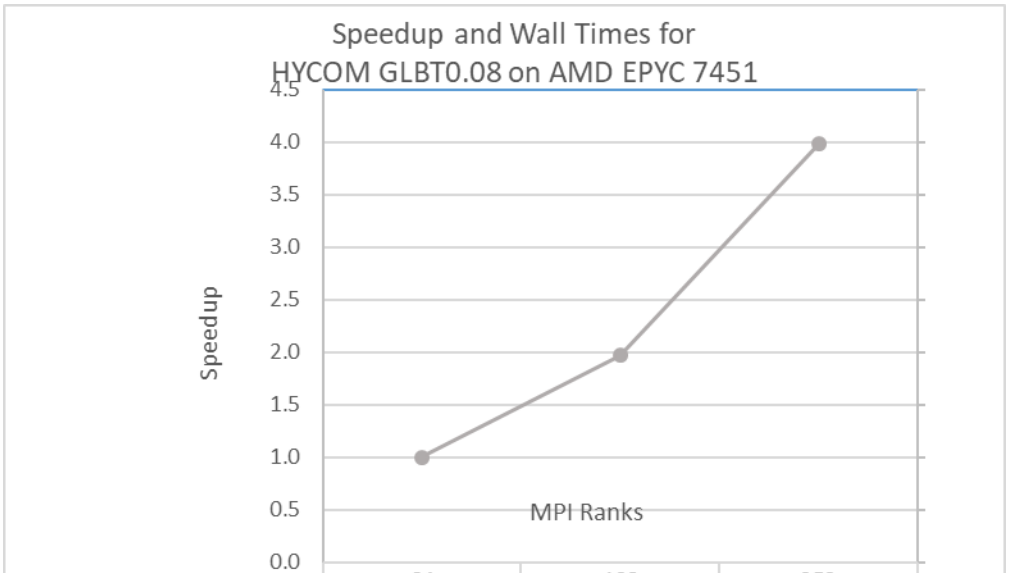
GLBT0.08 Using the EPYC 7451 Processor Cluster

We ran the larger test case, the GLBT0.08 benchmark, with the same AMD EPYC 7451 processor cluster (Cluster 1). Figure 4 shows the result, which is also five run averages where the speedup is relative to the lowest core count run (64-cores in this case).

The data also show an ideal scaling walltime, computed by scaling the initial walltime at 64 cores by the fractional increase in cores.[2]

As with the smaller test case, scaling trends in a linear fashion, but with changes in the slope which are due to cache effects, the discussion of which is beyond the scope of this paper.

Note that due to the size of this cluster (and the designated HYCOM rank counts mentioned previously), only three rank count steps could be run with Cluster 1's EPYC 7451 processors.



	64	123	252
Speedup	1.0	2.0	4.0
Walltime 7451	16728	8495	4206
Ideal Scaling Walltime	16728	8704	4248

Figure 3: Higher resolution test case GLBT0.08 using EPYC 7451 cluster

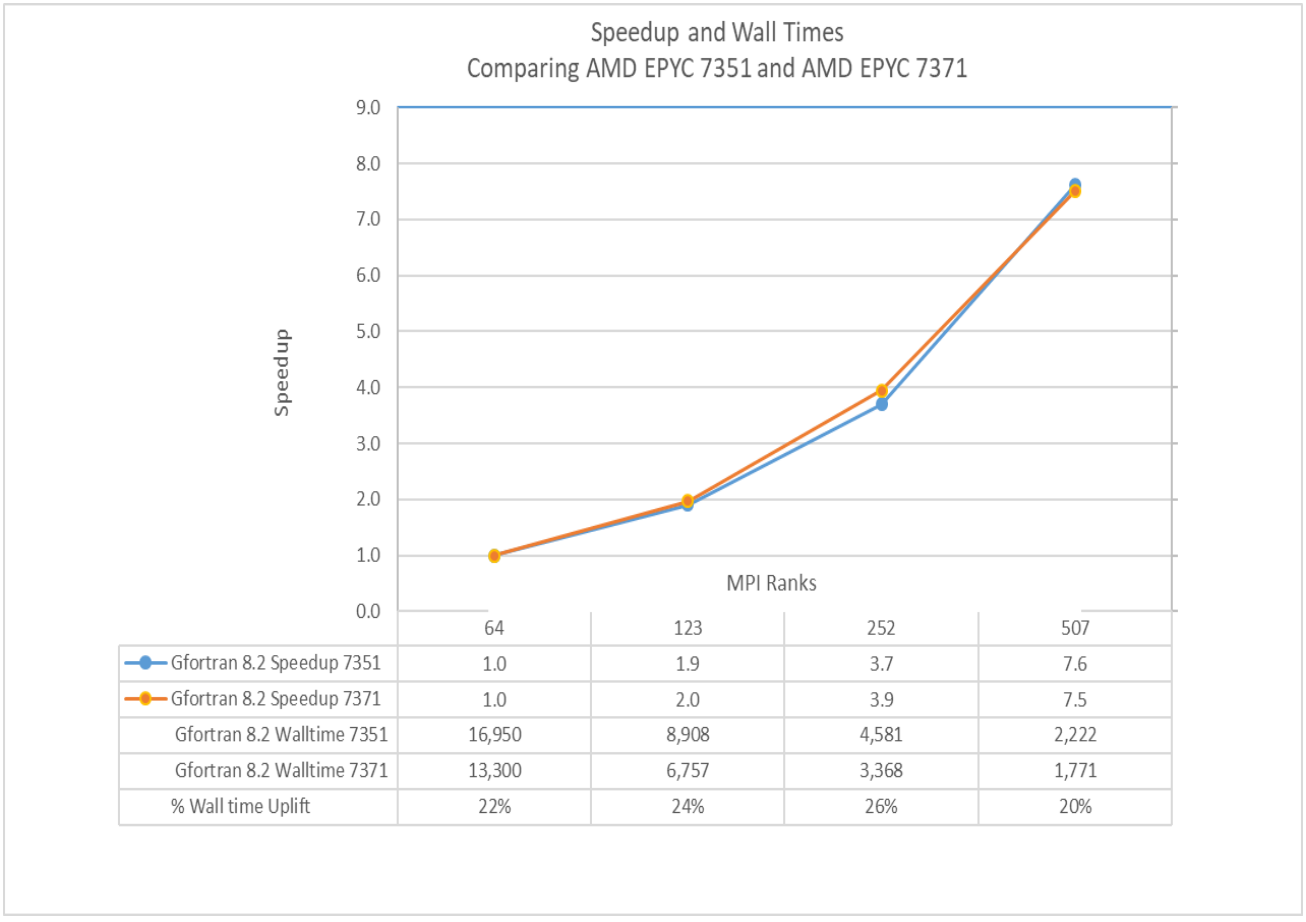
GLBT0.08 Using the EPYC 7351 and EPYC 7371 Processor Clusters

Next, we ran the larger GLBT0.08 benchmark on Cluster 2 and Cluster 3, each with the EPYC 7351 and EPYC 7371 processors as described in Table 2. These processors both have 16 cores (eight fewer than the 24-core EPYC 7451).

The EPYC 7351 has a base frequency comparable to that of the EPYC 7451 used previously, while the EPYC 7371 has a higher base frequency of 3.1 GHz.

All-core-boost frequency is also higher for the EPYC 7371 at 3.6 GHz, compared to 2.9 GHz for the EPYC 7351 and EPYC 7451.

The results of these benchmark runs are shown in Figure 5 along with comparative percentage uplift, i.e, percentage reduction in wall clock time. Note that the EPYC 7371's scaling behavior is difficult to distinguish, as it is nearly identical to the EPYC 7351 scaling behavior. The wall clock times are considerable less for the higher frequency EPYC 7371 processor.



	64	123	252	507
Gfortran 8.2 Speedup 7351	1.0	1.9	3.7	7.6
Gfortran 8.2 Speedup 7371	1.0	2.0	3.9	7.5
Gfortran 8.2 Walltime 7351	16,950	8,908	4,581	2,222
Gfortran 8.2 Walltime 7371	13,300	6,757	3,368	1,771
% Wall time Uplift	22%	24%	26%	20%

Figure 4: Scaling runs comparing speedups of the EPYC 7351 and EPYC 7371

Benchmark Results: Comparative Observations

Scaling of parallel workloads is affected by many factors beyond the processor: communications overhead, I/O bandwidth, memory bandwidth, network bandwidth, the number of nodes, and the file system, to name a few.

In contrast, the frequency of a CPU affects the number of instructions per second that it can execute. Therefore, any measured speedup in the system can be attributed to the CPU frequency only if there is enough bandwidth for all of these factors to feed the cores of the processor and prevent them from going idle.

There is clear evidence that performance uplift for these HYCOM test cases is frequency-sensitive. Table 3 shows the processors tested, their core counts, all core boost frequency, and both the frequency and wall clock time improvement, i.e., performance uplift, of the EPYC 7371 over the other two processors.

Figure 5 and Table 3 both show a performance uplift for the EPYC 7371 which is roughly proportional to its increase in frequency over the other processors. This performance uplift fluctuates between 20-26% over the EPYC 7351, and remains steady at 20% over the EPYC 7451.

Processor	Cores	All Core Boost Frequency (GHz)	EPYC 7371 Frequency Advantage	EPYC 7371 Performance Advantage
EPYC 7371	16	3.6	n/a	n/a
EPYC 7351	16	2.9	24%	20% - 26%
EPYC 7451	24	2.9	24%	20%

Table 3: Frequency vs. Uplift for all the processors tested.

It follows that HYCOM's performance has a frequency-dependent component, which the higher frequency EPYC 7371 processor is able to exploit.

Cores vs. Frequency

The other factor to consider is the number of cores per processor, and the additional machinery these may be able to bring to bear on the workload. To determine this, we compared the 24-core EPYC 7451 processor with the 16-core EPYC 7351 processor, both of which have the same all-core-boost frequency of 2.9 GHz.

The 24-core EPYC 7451 processor does in fact deliver better performance than the 16-core EPYC 7351 processor. In particular, comparison of the wall clock time for the 24-core EPYC 7451 versus the wall clock time for the 16-core EPYC 7351 processor shows a performance uplift ranging from 1%- 8% for the higher core processor (see Table 4).

Processor	Cores	All Core Boost Frequency (GHz)	EPYC 7451 Core Count Advantage	EPYC 7451 Performance Advantage
EPYC 7451	24	2.9	n/a	n/a
EPYC 7351	16	2.9	50%	1% - 8%

Table 4: Core count vs uplift for processors with the same all-core-boost frequency: EPYC 7351 and EPYC 7451.

This demonstrates that at a constant maximum frequency, the eight additional cores of the EPYC 7451 cannot provide the same performance uplift for these workloads as does the higher frequency of the EPYC 7371 processor. Having 50% more cores only yields at most an 8% uplift in performance for the EPYC 7451 over the EPYC 7351 – whereas, having a 20% higher all-core-boost frequency allows the EPYC 7371 to achieve a 20% performance uplift with respect to the EPYC 7451, and 20% to 26% performance uplift over the EPYC 7351.

Observations

We have already established that HYCOM is sensitive to frequency, as demonstrated by the higher frequency 16-core EPYC 7371 outperforming lower frequency processors.

Observing that the EPYC 7451 performs better than the EPYC 7351, but performs worse than the EPYC 7371, leads to the conclusion that the processor with more cores to do the work wins if frequency is held constant, but only as long as system bandwidth is not an issue.

The fact that the EPYC 7371 outperforms the EPYC 7451, despite having fewer cores, also allows us to conclude that a higher frequency processor can outperform both a lower frequency processor with more cores to do the work, as well as a lower frequency processor with the same number of cores, so long as there is system bandwidth available to keep the cores fed.

Scale Out Behavior

Figure 5 shows that as the system scales out, the curves describing the wall clock times for the various processors begin to converge, indicating that the common factors that make up the overall system bandwidth become the dominant factor affecting performance as the benchmark test grows larger. This has important implications when choosing which processor to use based on the size of the workload.

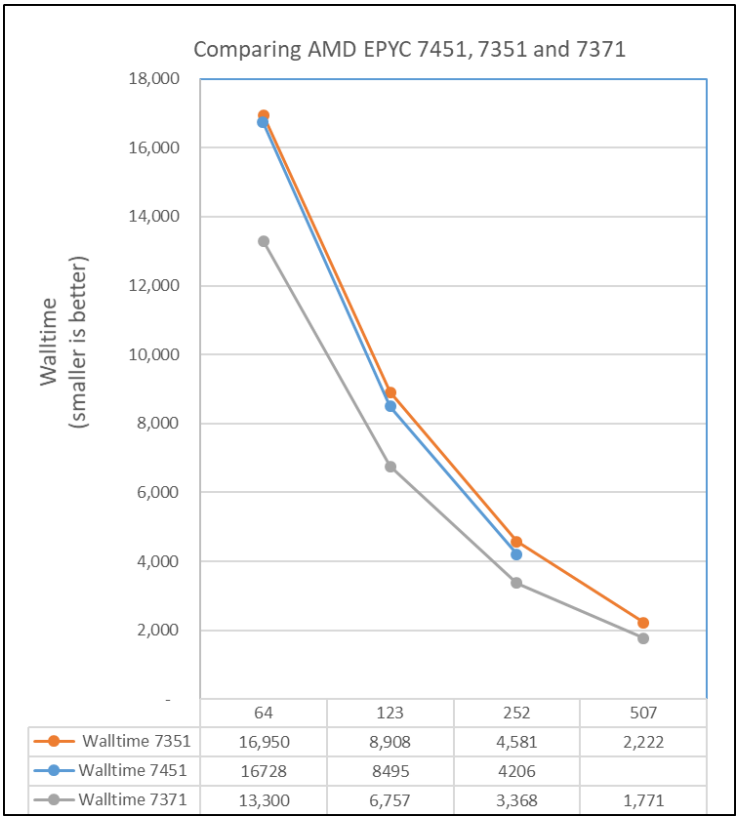


Figure 5: Overall walltimes for all processors tested: AMD EPYC 7451, AMD EPYC 7351, and AMD EPYC 7371

Conclusion

Scale-out testing of HYCOM global oceanic weather models on AMD EPYC hardware clusters showed linear scaling to the limit of available cluster resources on all tested processors.

The data presented should help in making decisions regarding how best to optimize TCO depending on software licensing models (cores vs. sockets vs. nodes).

Specifically, comparative testing of the high frequency 16-core AMD EPYC 7371 showed a 20% - 26% reduction in wall clock time as compared to the lower frequency 16-core AMD EPYC 7351 processor, and a steady 20% wall clock time reduction when compared to the 24-core AMD EPYC 7451 processor.

This last finding confirms that system bandwidth is an important factor for HYCOM performance, while frequency and core-count can vary based on the size of the workload.

AMD EPYC has been designed from the ground up for a new generation of solutions like HYCOM, with the entire feature set of the processor available regardless of the number of cores, along with ample I/O and memory bandwidth to help customers right-size their hardware to their specific needs.

Footnotes

1. This image taken from https://www7320.nrlssc.navy.mil/GLBhycomcice1-12_mnsd/skill.html, courtesy of the US Navy.
2. Ideally, we would expect the walltime to drop by the fractional increase in cores used – twice as many cores, half the walltime. Consider the walltime then at 64 cores, which was measured as 16728 seconds. Scaling by the core increase, we have (16728 seconds)*(64 cores /123 cores) = 8704 seconds, as shown. However, the walltime dropped by even more, to 8495 seconds. This superlinear speedup is a well-known cache effect (cf. Ristov, Sasko & Prodan, Radu & Gusev, Marjan & Skala, Karolj. (2016). Superlinear Speedup in HPC Systems: why and when?. (Researchgate))

DISCLAIMER