THE MEDIA TEAM

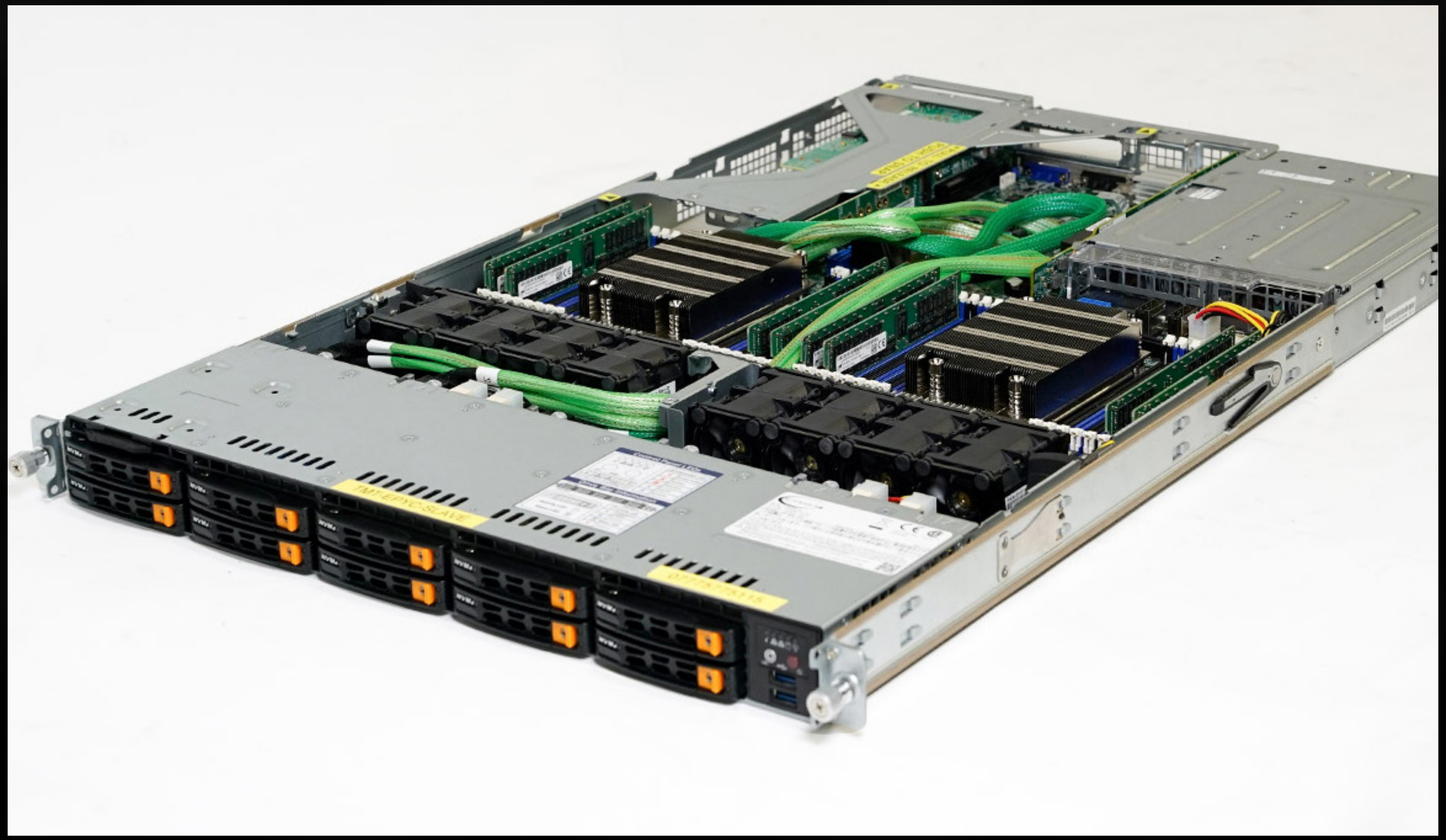
ADOPTING AMD EPYC

A Case Study

INTRODUCTION

Since 1998, The Media Team has been publishing an ever-growing number of eclectic websites specialising in independent technology news and reviews, modding, creative creation, food, and gossip. Working with brands across these segments, The Media Team unique selling points include bespoke content creation, in-house design, editorial partnerships and providing key insights through data analysis.

It has been identified that the IT infrastructure required to support the increasing traffic load from the various websites and projects has been under strain for a while, leading to slow server responses at times of peak load. Furthermore, whilst the underlying server(s) have been upgraded over time as new websites have been brought into the fold, The Media Team realises that a fundamentally new server solution is required – one that can comfortably deal with today's demands and greater load expected tomorrow.



SUMMARY OF EXPERIENCED CHALLENGES

The current Xeon-based server, comprising a dual-socket Intel Xeon E5520 offering a total of eight cores and 16 threads alongside 96GB of RAM, hosted at a local datacenter, and upgraded over time, is found wanting in a number of ways. These are categorised as follows:

- Significant slowdowns in high-demand webserving caused by older server processors hitting peak utilisation too quickly. These cannot be readily upgraded without changing platforms.
- A lack of memory bandwidth and capacity that inhibits performance and queries on large-scale databases. This is particularly evident as the load exhausts main memory and runs from a much slower swapfile.
- Older storage standards prohibit the use of newer, faster technology. PCIe 4.0, only available on AMD platforms, enables twice the bandwidth as PCIe 3.0 at the same link width.
- The current server has poor energy efficiency compared to newer solutions, impacting upon TCO.
- Newer servers offer significant upgrade potential in the same form factor as the present solution

Keeping to the same platform, however, enables The Media Team to upgrade the extant infrastructure to a dual-socket Intel Xeon X5675 offering a total of 12 cores, 24 threads, a higher peak turbo speed, and up to 196GB of RAM. Even so, whilst performance improves with respect to requests per second (rps) and serving latency, there's limited scope to expand further.

PROPOSED SOLUTIONS

Understanding that the current server setup cannot cope with the demands placed upon it, even if upgraded to its maximum specification, The Media Team needs a server solution that can scale as the company grows, offering up to 10x the current requests per second (rps) and significantly reduced page-serving latency during high-profile product launches and hot stories. Achieving this huge increase in serving potential can be realised one of two ways: firstly, the server and application requirements can be ported over entirely to established providers such as Amazon Web Services (AWS), Digital Ocean, or Linode, scaling as demand escalates.

The other option is to build a brand-new, scalable server platform to The Media Team's specifications, addressing the key challenges identified above. In particular, the platform needs to offer far superior performance, greater upgradeability, holistic platform advantages, reduced total cost of ownership (TCO), and general productivity and flexible deployment and virtualisation improvements over the incumbent solution. It also needs to offer more value than comparable cloud platforms.

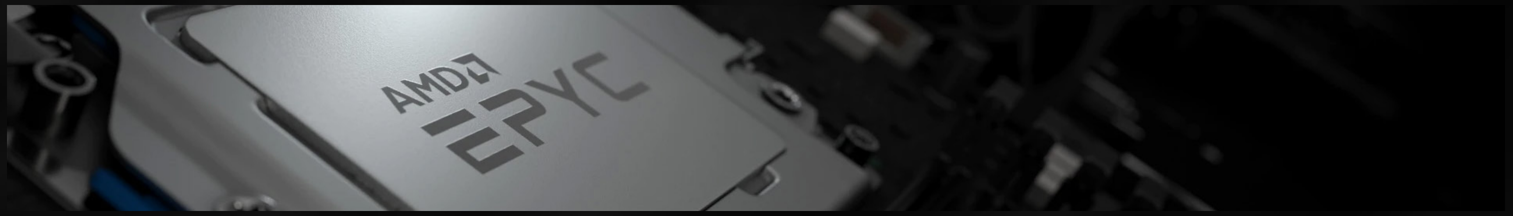
"...The Media Team needs a server solution that can scale as the company grows."

Unrivalled power in a 2U chassis; easier manageability and simpler upgrading

THE MEDIA TEAM

CHOOSING AMD EPYC

After careful deliberation of the requirements and evaluation of the cutting-edge server solutions available from Intel and AMD, both of which are solid candidates for upgrading the present setup, The Media Team chose to implement a dual-socket AMD 2nd Generation EPYC 7662 server - 128 cores, 256 threads, 512GB RAM - as it clearly demonstrated leadership in outright performance – especially in virtualisation instances - platform-level features, future upgradeability, and reduced TCO compared to an Intel Xeon alternative.



There's more to it than pure performance alone, however. AMD EPYC offers a balanced solution to running virtual machines, with each serviced by ample connectivity, storage and memory size, speed and capacity. Furthermore, with security being of such rightful concern, AMD's Secure Encrypted Virtualization (SEV) technology, hardware baked into each processor, offers peace of mind as it stops the contents of one virtual machine being read by another VM through transparently encrypting the memory of each VM with a unique key, all with minimal performance overhead. As The Media Team plans to run multiple websites across VMs on one server, where uptime is paramount, knowing that VMs are cryptographically isolated from one another is of great importance.

That's not all, either, as any solution adopted by The Media Team needs to be at the forefront in every area - outright performance, connectivity, scalability, energy efficiency, and security. Considering the latter angle in a wider context, clean-sheet architectures such as AMD EPYC are better positioned to withstand and mitigate the growing number of security vulnerabilities that exploit older hardware. To this end, AMD EPYC's architecture, which includes a secure processor for encrypting system memory, has inherently more robust defences against side-channel attacks than Xeon hardware.

Running benchmarks that replicate webserving performance compared to the incumbent solution shows:

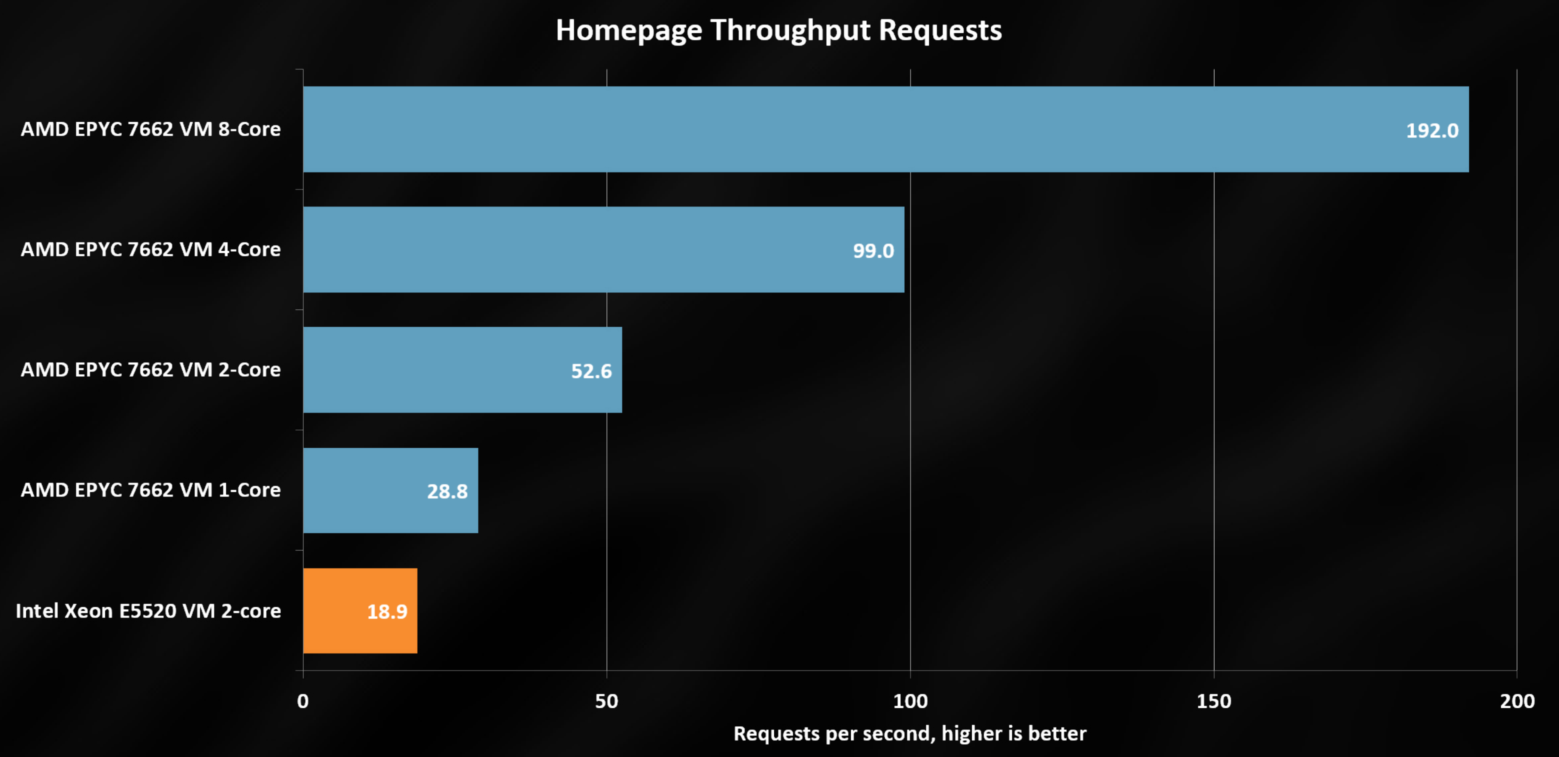
- 10.2x greater performance in a virtualised environment.
- 96 percent reduction in homepage request latency.
- 8U rack space reduced to merely 2U.
- Ability to upgrade to next-generation CPUs without changing the underlying infrastructure.
- Ability to run super-fast storage with unfettered access to the CPUs.

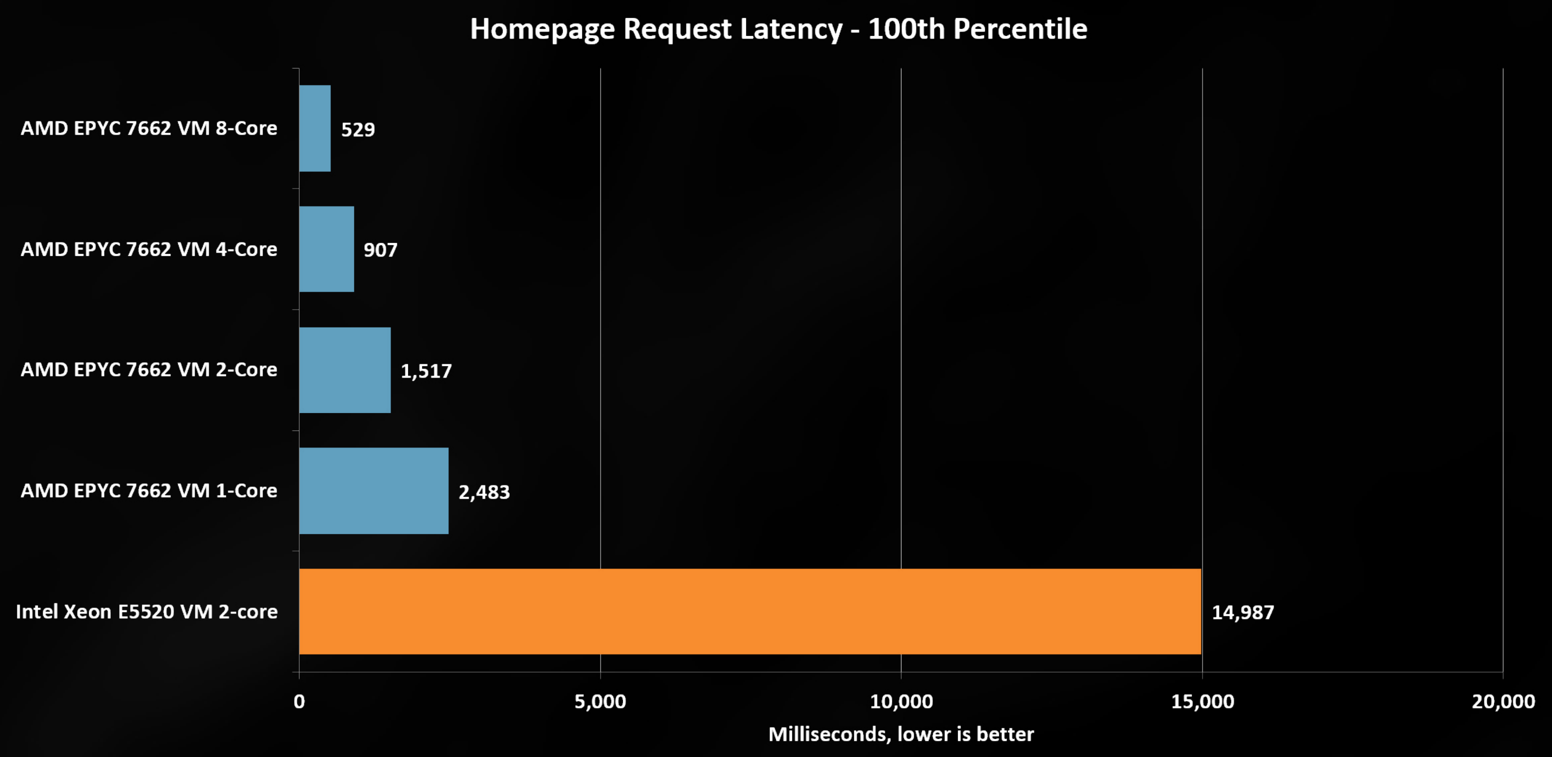
It is this impressive computational density that makes AMD EPYC the only viable choice for a much-needed full-server upgrade for The Media Team. Reducing the complexity of the server setup to a single 2U solution also has the positive knock-on effect of ensuring manageability is easier, upgrading is simpler, and general administration and server maintenance considerably more straightforward.

The Boston-built 2U rackmount server

PERFORMANCE

Putting the performance advances into stark context, The Media Team benchmarked the homepage request throughput and access latency at various percentiles, underscoring huge increases on both fronts. Benchmarks were conducted using the popular ApacheBench software to bombard the servers with requests. They were run from a virtual machine on the same network which had access to 64 cores and 128GB of RAM. The application tested was run using 4GB of RAM in all tests using two cores on the old platform and 1-8 cores on the new platform. Tests were run using 10,000 requests with a concurrency of 64 in apache bench



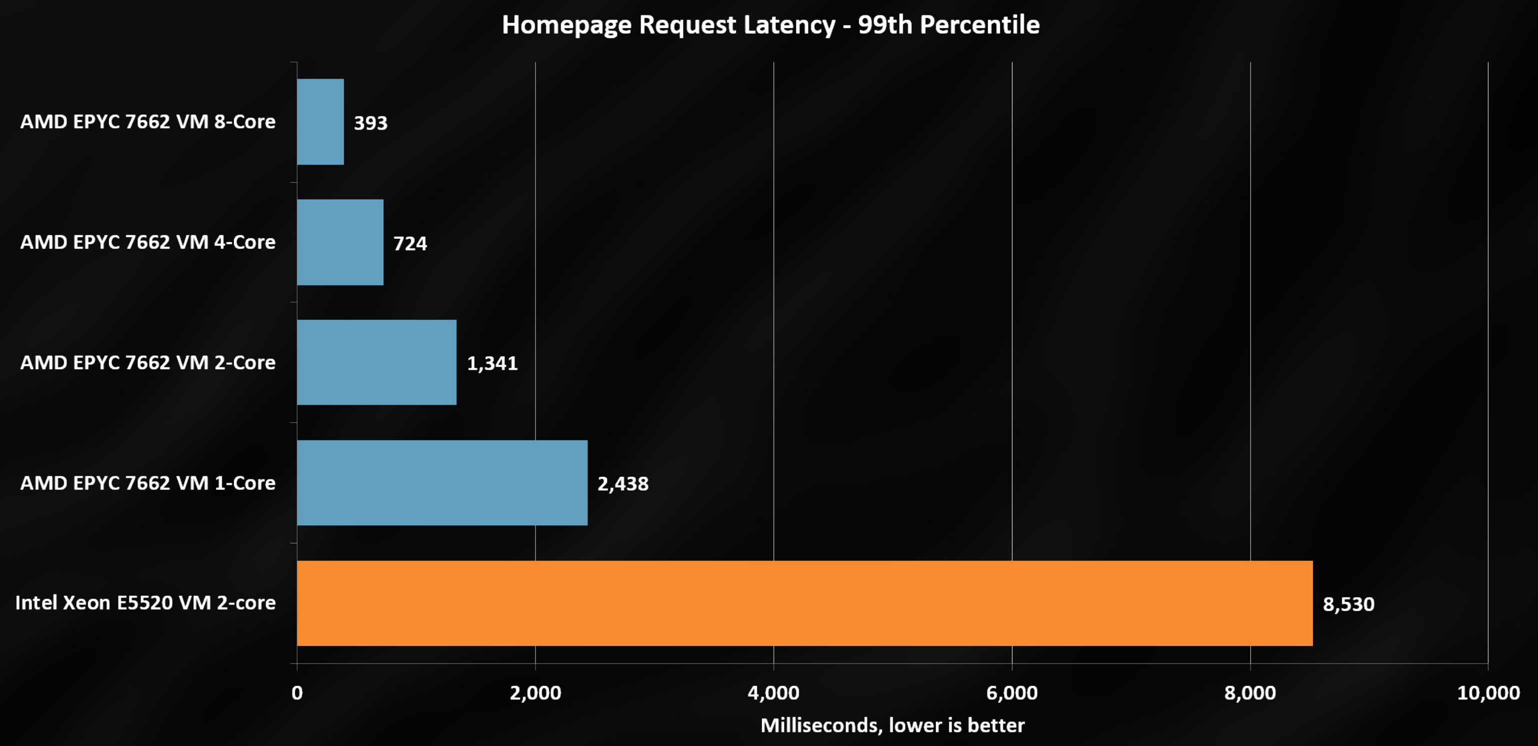


The above graph shows performance when running The Media Team's Cranble website using the methodology described above. Each AMD EPYC 7662 has 64 cores at its disposal. Even so, compared to the incumbent solution, requests per second increase from 18.9 to 52.6 - a 2.78x improvement - when evaluated over the same number of cores, ably demonstrating the significant advances in core microarchitecture run on a cutting-edge, scalable platform. Increasing the core count to eight improves performance by over 10x whilst still leaving plenty of scope for development on the same platform.

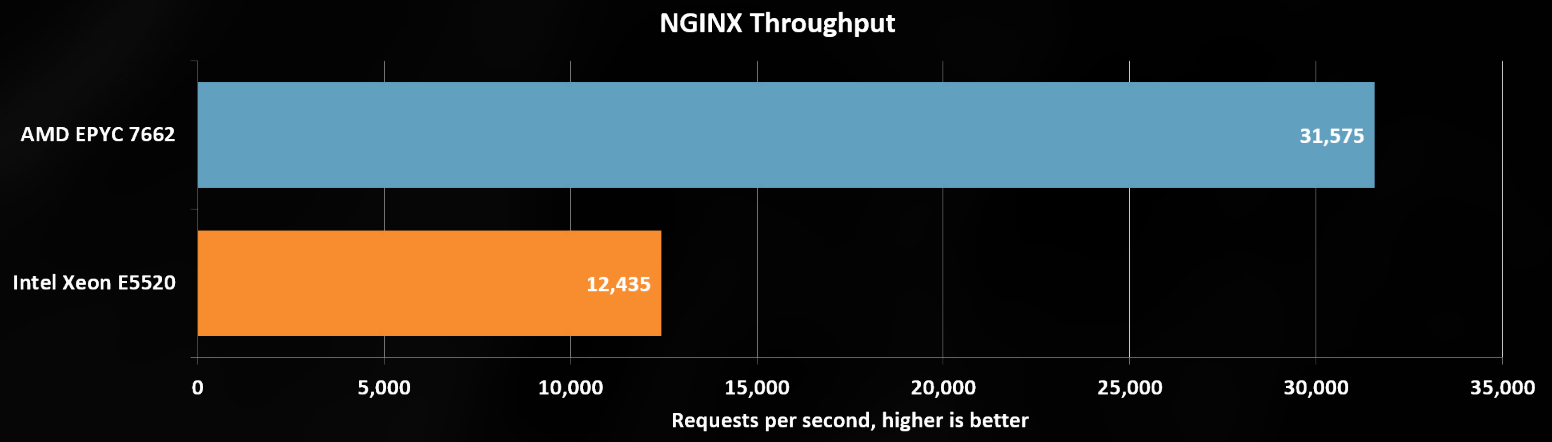
Arguably just as important as the number of requests is the latency of serving the page(s). Having every one delivered to the readers on the existing solution, limited to two cores for benchmarking purposes, takes up to 15 seconds, which is clearly unacceptable. That time drops to 1.5 seconds for the equivalent number of cores for the AMD EPYC 7662, reducing to just 0.5s with only eight cores. In other words, like for like, there's a near-3x request throughput improvement while all pages are served in 1/10th of the time.

THE MEDIA TEAM

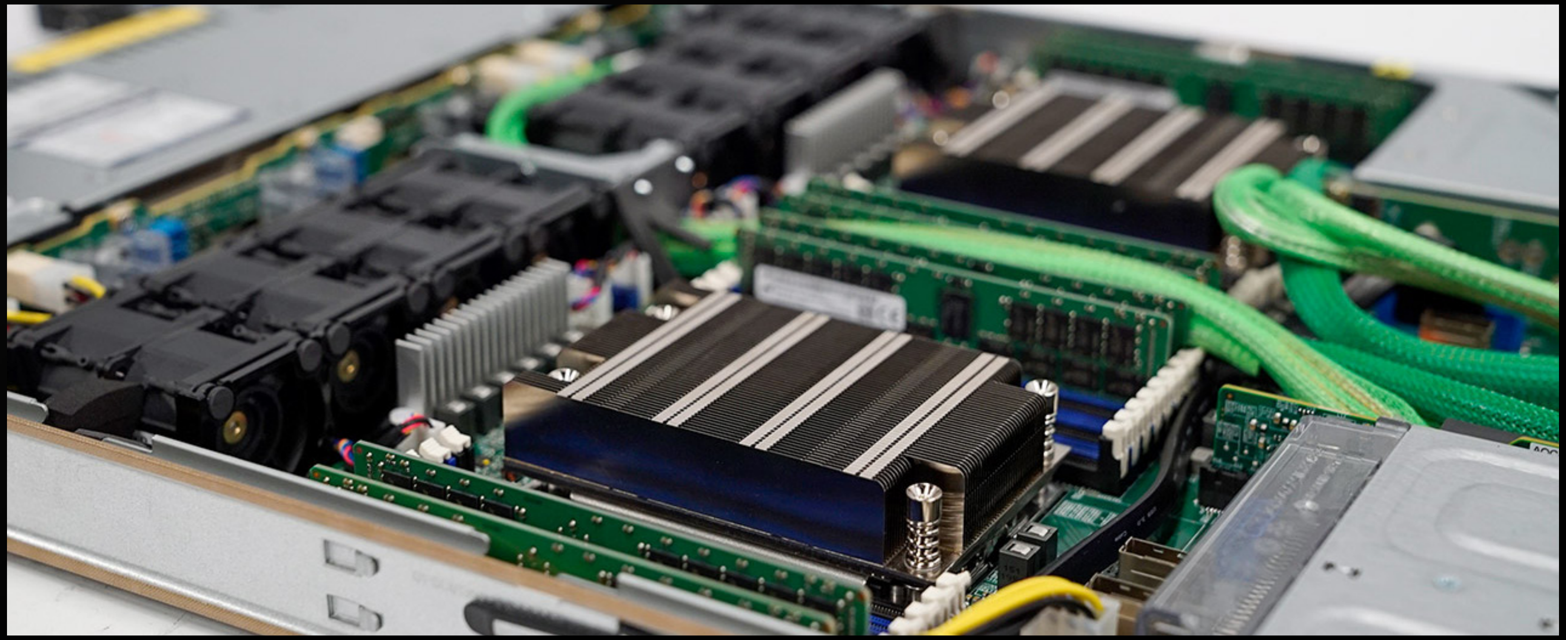
PERFORMANCE



Dropping down to the 95th percentile, meaning the performance level most will experience, still shows a marked improvement for AMD EPYC. Of course, a newer Intel solution is absolutely going to be much closer to the AMD EPYC performance, though our recent testing for a dual-socket 7742-based server shows that AMD holds a commanding lead against an Intel dual-socket Xeon Platinum 8280 in CPU-intensive tasks that ably mirror the kinds of loads imposed by The Media Team's servers.



NGINX is open source software for web serving, reverse proxying, caching, load balancing, media streaming, and is great for webserving and application delivery. Performance for EPYC, as expected, stands head and shoulders above a decade-old Xeon server. As many popular sites use NGINX as a host, high throughput is indicative of a smooth browsing experience for the end-user.



It's more beneficial for The Media Team to have an on-prem deployment

COST AND POWER

The performance of AMD's EPYC 7662 2P solution is undeniably strong and consistent because all computation is contained within one server, minimising rackspace and complexity of balancing between older, slower servers. In terms of cost, this EPYC installation costs £48,400 in year one, rising to a total of £65,000 in year three, compared with an equivalent Amazon Web Services (c6g.16xlarge, plus 14TB of Elastic Black Storage) three-year reservation costing £7,526 per month, or £270,000 over the same three-year period. It is over four times cheaper for The Media Team to go down this route.

The installation cost of a 2P Intel Xeon 8280 is approximately £7,000 higher for the matching components within the server, even though, as benchmarks have previously shown not only is absolute performance substantially lower, there's a nebulous upgrade path for the next three years. Compare that with AMD's socket-compatibility promise for next-generation EPYC processors offering even more performance.

> **"...It is over four times cheaper for The Media Team to go down this route."**

Cost is only one part of the equation. Being able to harness as as much performance, efficiently, is another. We can compare this by looking at the 2U server performance in Cinebench R20. The AMD EPYC 7662 solution, comprising two processors, returns a score of 18,203 whilst consuming 621W - or 29.3 marks per watt. Meanwhile, the Intel solution it replaces, also presented in a 2U form factor, also comprises two Xeon 5520 processors that return a score of 1,428 marks at a mains power consumption of 362W - or 3.9 marks per watt. Architecture and process advancements over the years leads to a 7.5 performance-per-watt improvement when comparing the same number of processors.

STORAGE SOLUTIONS

Request performance is the key criterion by which The Media Team evaluates new server hardware. However, manageability, redundancy and outright performance of the storage subsystem is another key area ripe for a makeover in conjunction with general server updates. AMD EPYC's ability to offer an unprecedented number of high-speed expansion lanes and flexible storage support means that The Media Team can also invest in a bespoke storage setup that achieves the keenest balance between the three aforementioned factors.

> **"...Like the AMD EPYC hardware, there's significant upgrade potential as and when needed."**

To this end, a RAID10 array provides the optimum balance between performance and redundancy for general storage while RAID1 makes most sense for the boot drives. In particular, a Broadcom MegaRAID 9460-8i card is used in RAID1 for two Intel Optane SSD DC P4801X Series 100GB drives hosting the operating system. Meanwhile, a Broadcom MegaRAID 9560-16i card connects four Micron 9300 Pro 3.84TB 2.5in NVMe U.2 SSDs in RAID10, providing 7TB of usable storage for the virtual machines and 100GB for the VMWare ESXi install. Appreciating the capabilities and range of the Broadcom controllers, the storage subsystem can easily be expanded over time to more drives with larger capacities. What's more, if the demand arises, a Broadcom PCIe 4.0 controller can also be installed for higher bandwidth and performance still.

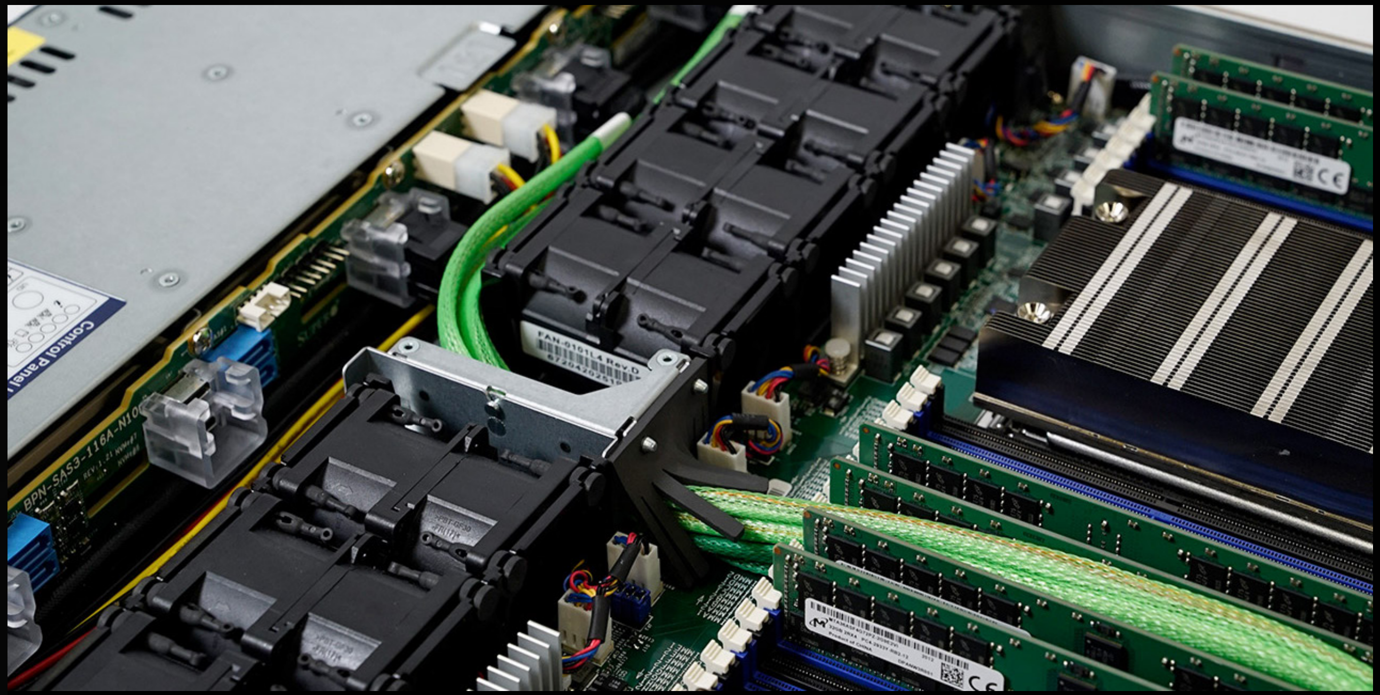
CONCLUSION

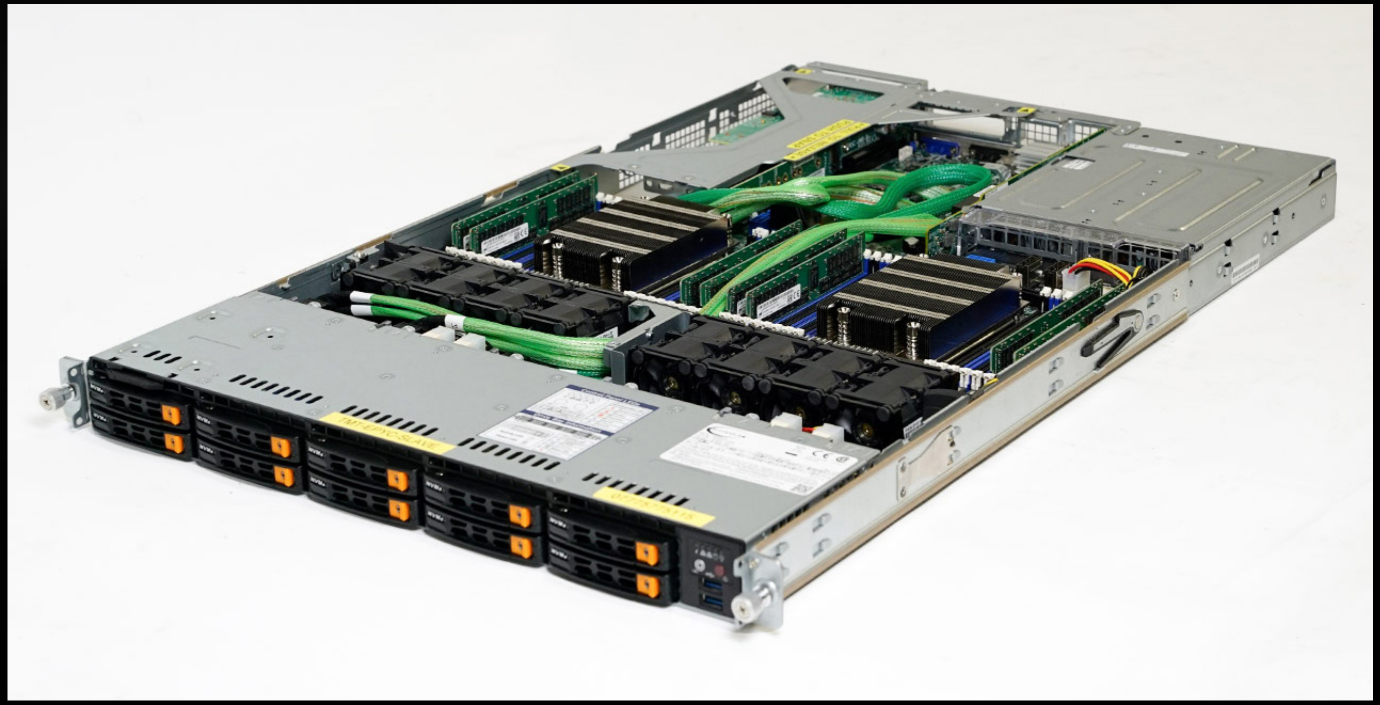
This case study ably demonstrates that for a small technology business such as The Media Team, specialising in online content serving and in-house design, an AMD EPYC installation represents the preferred combination of class-leading performance, consistent upgrade path, and lower total cost of ownership than going for a broadly equivalent Intel Xeon alternative or to a one-stop provider such as AWS.

The Media Team has grown over time by serving interesting, rich, quality online content across a number of genres. A key part of delivering an ever-growing number of pages in a timely fashion is the underlying server architecture. As a small business, any relatively large server infrastructure investment needs to provide long-term service in an easy-to-manage package, all without breaking the bank.

Seen through the lens of recent high-traffic acquisitions that overload the incumbent setup, The Media Team had to invest in a grounds-up server upgrade that would provide ample performance for today and tomorrow, great scalability and upgradeability, class-leading virtualisation for the numerous websites, cutting-edge security, impressive density and power consumption, and fantastic density all in value-centric package that can be managed by the company's only IT manager.

After deliberation it became clear that an on-prem AMD's EPYC server infrastructure excels in each required metric whilst reducing hardware complexity and offering clear futureproofing as and when The Media Team server needs grow.





For the server needs of today and tomorrow