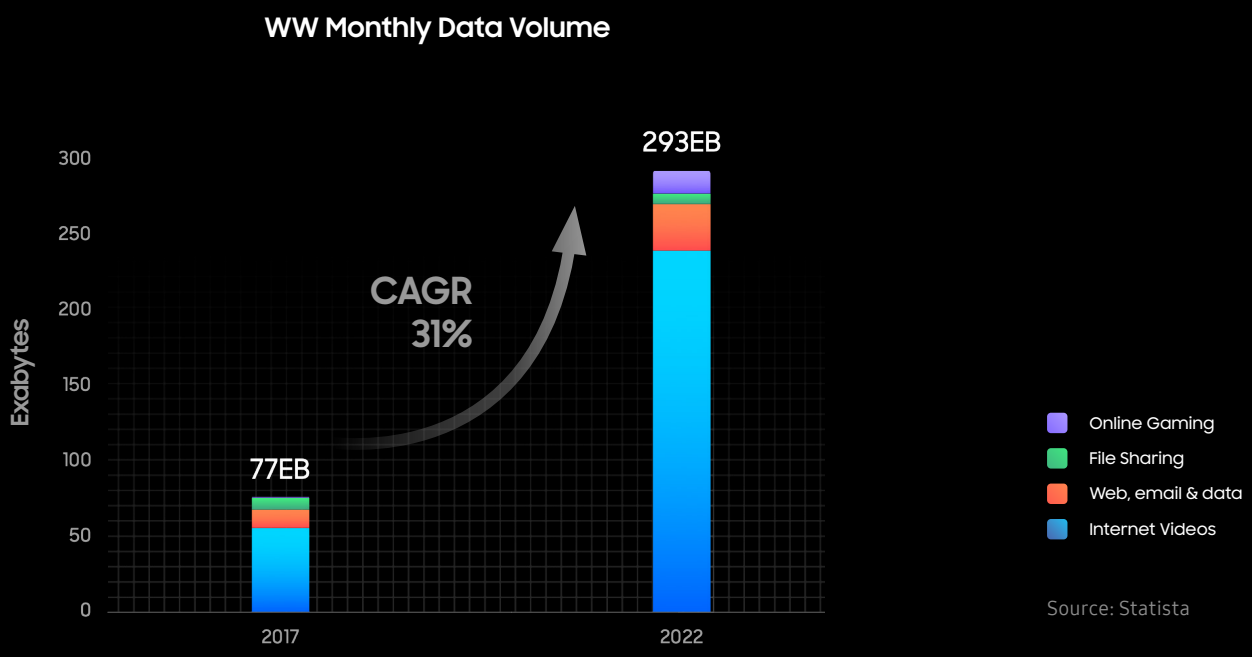
SAMSUNG | AMPERE.

Revolutionizing Data Processing to Support Cloud Native Workloads

Case Study

SAMSUNG

There's a revolution occurring in the cloud compute space—and it's all about achieving performance, scalability and power efficiency. To provide enough horsepower for the rich content, data analytics, and AI and machine learning that will drive future innovation and progress, a completely new hardware paradigm that drives performance while being efficient and sustainable. By combining state-of-the-art, power-efficient cloud native CPU, memory and storage solutions, Samsung and Ampere are collaborating to make it happen.



The amount of data generated in today's digitally focused world is astonishing. By 2025, consumers are expected to generate 175 trillion gigabytes[1] of new data per year. The more data we have, the more we can apply it to developing innovative products and services, boosting efficiency and cost-savings in business and operations, and increasing our knowledge about the world around us.

As organizations collect, process and store increasingly vast volumes of data, cloud computing takes center stage. According to Gartner, the global cloud revenue will total $474 billion in 2022, up from $408 billion in 2021, and more than 85% of organizations will embrace a cloud-first principle by 2025.[2] In fact, Gartner predicts organizations won't be able to fully execute on their digital strategies without using cloud-native architectures and technologies.

1. International Data Corporation (IDC) predicts stored data will grow by 61% by 2025, totaling 175 zettabytes.
2. Gartner predicts that in 2022, global cloud revenue will total $474 billion, up from $408 billion in 2021 and could surpass non-cloud revenue for relevant enterprise IT markets in the next few years.

Out with the Old–In with Cloud Native

Cloud native applications are designed to scale out efficiently and take advantage of a massive number of highly efficient cores in the data center vs legacy applications which are monolithic.

While potential applications for cloud computing seem limitless, the infrastructure that supports them must be able to handle cloud native workloads unlike legacy hardware designed for client-server architectures of the past. Today's cloud infrastructure is challenged to address the key requirements of modern cloud computing:

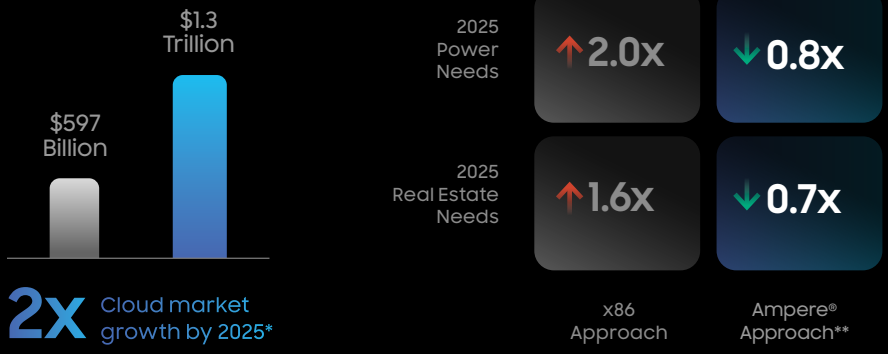
> Data centers require energy capacity of close to 100 MW—that's equivalent power for about 80,000 U.S. homes and adds up to about 1% of the world's electricity demands.
>
> **Greenpeace**
> Global Environmental Group

Performance
People expect instantaneous responses from cloud-based applications. Imagine running a search on Google and having to wait even five seconds for results! For AI-driven services related to sales and revenue generation—think customer service chatbots—a slow response could mean lost dollars and damaged reputation.

Scalability
One of the main benefits of offering cloud-based services is the ability to scale along with demand. Such scalability can't be dependent on adding servers to the infrastructure—it must be automatic and instantaneous. Whether you have one user or hundreds of users connecting to your server, you must be able to deliver the same level of performance and service continuity, without hesitation.

Sustainability
According to Greenpeace, data centers require energy capacity of close to 100 MW—that's equivalent power for about 80,000 U.S. homes[3] and adds up to about 1% of the world's electricity demands. But 84% of executives[4] think sustainability initiatives are necessary to be competitive today or in the future. Power-hungry processors work against sustainability initiatives and drive up overhead, and they limit the number of users who can be served simultaneously without building new facilities and servers.

3. Greenpeace reports that Apple's $1 Billion iData Center in North Carolina will consume as much as 100 MW of electricity, and opens this Spring.
4. According to IMD, 62% of executives consider a sustainability strategy necessary to be competitive today, and another 22% think it will be in the future.



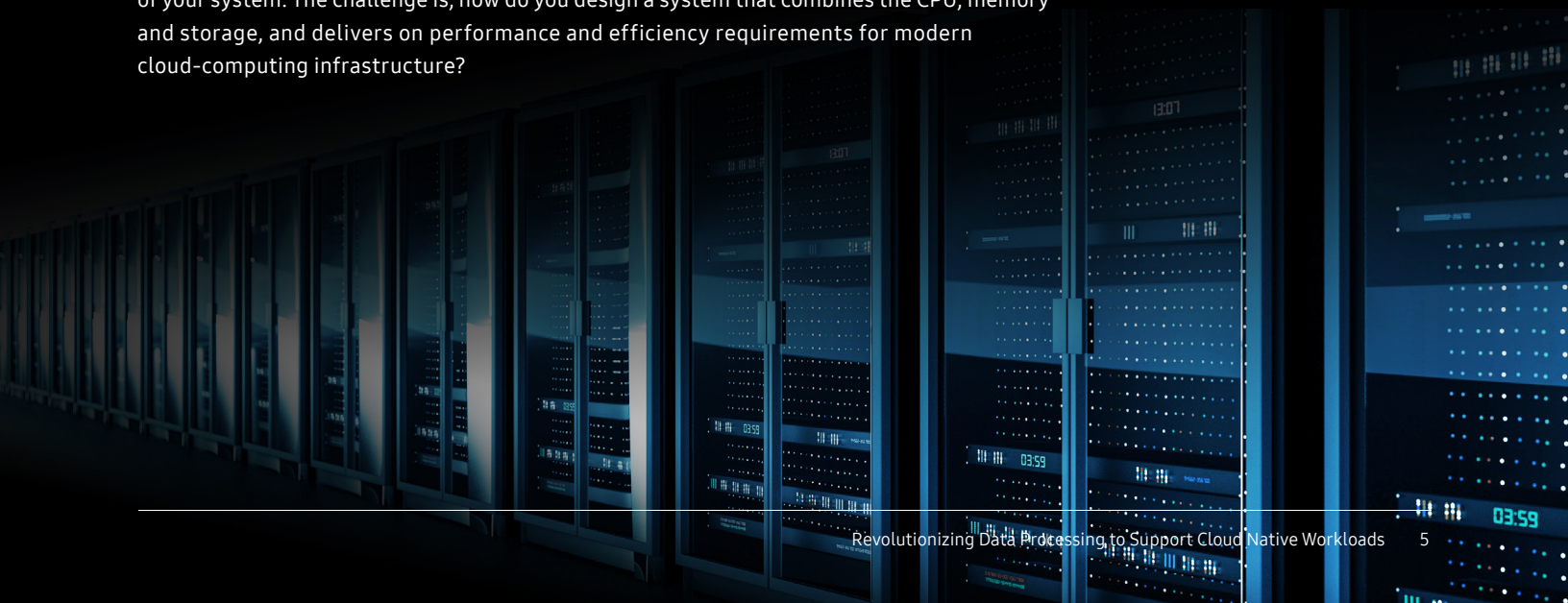
*Source: Worldwide Foundational Cloud Services Forecast 2021–2025—The Base of the Digital-First Economy (IDC #US46058920

**Used Ampere® internal modes and analysis to identify total compute demand, power consumption numbers, and real estate footprint for x86 and Ampere processors in 2025

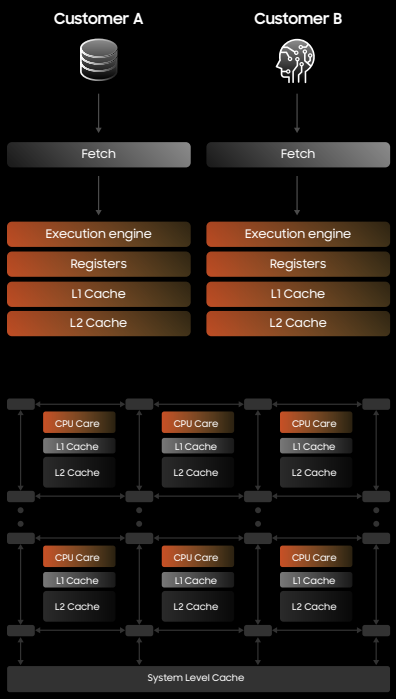
Over the past decade, x86 processors have evolved to provide increased performance, but at the expense of power efficiency. On the flip-side, power-efficient processors such as those used in smartphones and more compact computing devices cannot provide the scale or performance necessary for cloud native workloads. For cloud we need performance and efficiency at the same time. Adding to the challenge is the need for increased capacity—cloud applications and workloads are extremely data-intensive. However, memory and storage—the two essential components for delivering this capacity—can drive up the total costs of running a data center, if they're not designed with power efficiency in mind.

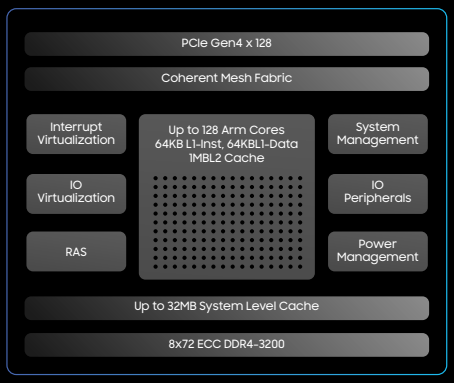
For high-performance core processors to perform as expected when processing large data models, datasets and 4K+ video, increasingly fast DDR speeds are essential—but processors must enable those speeds without driving up power consumption and costs. In most servers, 80%+ of the power consumption comes from the CPU; but in modern server environments—some of which have 16 channels across multiple sockets and 32 DIMMs or more—memory begins to contribute substantially to power consumption.

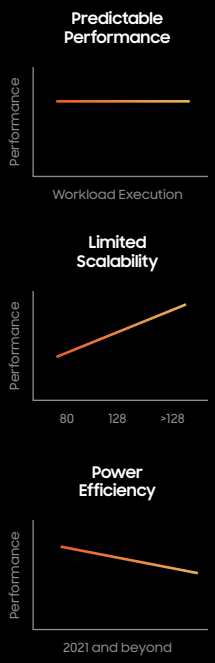
Storage performance is also critical for handling data-intensive workloads, and depending on the configuration—you may have high-performance storage for primary compute nodes and high-capacity storage on the back-end—storage can also add to the total power consumption of your system. The challenge is, how do you design a system that combines the CPU, memory and storage, and delivers on performance and efficiency requirements for modern cloud-computing infrastructure?



Unmatched Power & Efficiency for the Cloud-Compute Landscape







Ampere is re-defining the future of hyperscale cloud and edge computing with the world's first Cloud Native Processor, the Ampere® Altra® Processor family. These processors are complete system-on-chip (SOC) solutions that deliver the performance, scalability and sustainability necessary for modern cloud computing environments. With up to 128 cores, Ampere's 64-bit processors offer predictable high performance, scalability and power efficiency, and lend themselves to high utilization to support very large modern cloud native edge computing workloads and applications. Combining them with Samsung memory and storage solutions results in a powerful solution for the cloud.

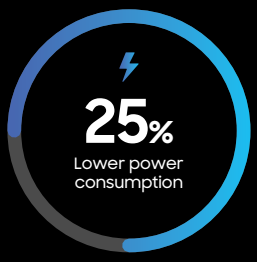
"Ampere's processors deliver a new standard to the hyperscale market, in which high performance processing with reduced power consumption and scalability can save organizations millions and improve profits," said Madhu Rangarajan, Vice President of Products at Ampere. "Samsung's memory and PCIe SSD storage technology complements our cloud native CPUs by providing unmatched power efficiency and reliability—enabling us to develop cutting-edge server systems that are poised to transform the cloud compute landscape."

> With Samsung's memory solutions validated extensively with Ampere CPUs, we have peace of mind that such errors [uncorrectable memory errors] won't occur.

Madhu Rangarajan
Vice President of Products, Ampere

Ampere leverages Samsung's DDR4 RDIMM/LRDIMM memory, an optimized solution for enterprise applications that provides a highly virtualized environment and high-performance computing. With the latest 10nm class manufacturing technology, a monolithic 16Gb die solution and 3DS TSV DRAM package technology, Samsung's DDR4 RDIMM/LRDIMM enables a wide range of density configurations with lower power consumption than alternative high-density solutions, reducing overall TCO in the data center.

Samsung's DDR4 RDIMM/LRDIMM delivers up to 3,200 Mbps for easily processing massive workloads with unprecedented speed, and 4 bank groups (16 banks), reducing interleaving delays along with 1TB/s system memory. Importantly, the solution leverages Samsung's industry-first 1xNM process technology, along with 1.2V low operating voltage and a Pseudo Open Drain (POD) interface, reducing power consumption by up to 25% while boosting performance.



Less energy, greater efficiency.

Samsung continues to collaborate closely with Ampere to validate the newest D1z nm 8Gb and 16Gb DDR4 solutions—the most power-efficient solutions in the data center market today.

Importantly, Samsung's DDR4 solutions are extremely reliable. "Reliability is critical to cloud environments," said Rangarajan. "In dense core implementations such as those enabled by our cloud native CPUs, you may have 256 users running virtual machines across 256 cores, and uncorrectable memory errors can bring the whole system down and impact hundreds of users," he said. "With Samsung's memory solutions validated extensively with Ampere CPUs, we have peace of mind that such errors won't occur."

Close Collaboration Paves the Way for Innovation

According to Rangarajan, Ampere has collaborated closely and extensively with Samsung to ensure alignment across the companies' technology roadmaps. "We're working with Samsung to ensure our technologies intercept, work well together, and unlock new capacity points to support the various core and channel configurations our customers' request," he said. As memory can be one of the most expensive components in any platforms, ensuring the systems ship with the correct memory capacity saves on overprovisioning. "Samsung has been instrumental in helping us provide flexibility in delivering right-sized memory models to meet customer requirements," Rangarajan said.

SAMSUNG AMPERE

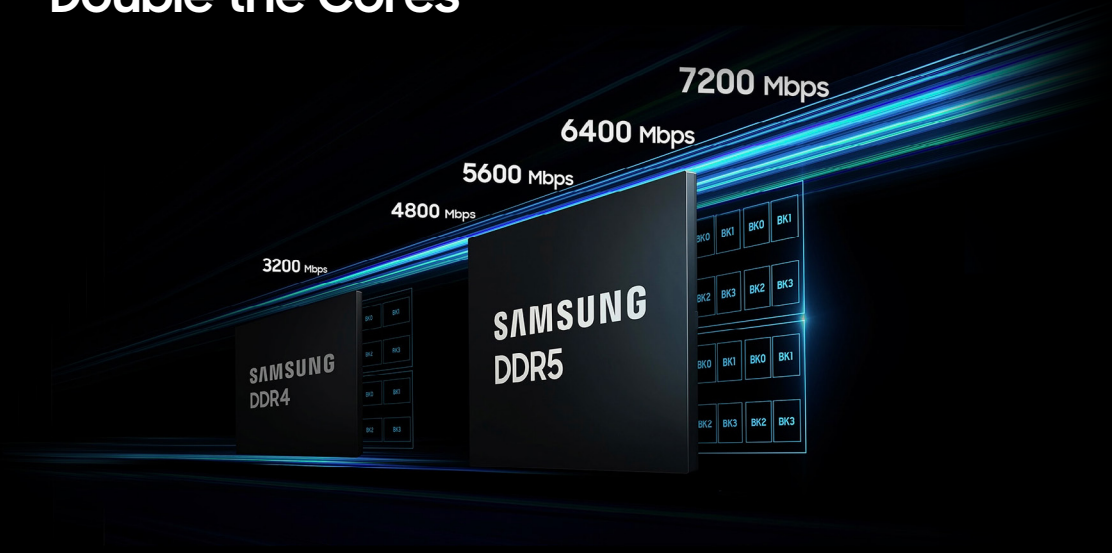
> "In our efforts to drive down server power consumption with innovative solutions, collaboration with Samsung will be essential."

Madhu Rangarajan
Vice President of Products, Ampere

Rangarajan added that memory is an area that's ripe for innovation, and Samsung is on the cutting-edge. "In our efforts to drive down server power consumption with innovative solutions, collaboration with Samsung will be essential," he said.

But it's not just collaboration and product synergy that puts Samsung's memory products on Ampere's bill of materials – their customers demand it. "Our customers have preferred vendors, and more often than not, Samsung is an important part of their list for memory and storage," Rangarajan said.

DDR5: Twice the Bandwidth, Double the Cores



Ampere is already evaluating Samsung's newly introduced DDR5 memory solutions, which provide the next level of speed, capacity and performance to meet the demands of industries experiencing tremendous bursts in data. Incredibly power-efficient and delivering exceptional transfer speeds of up to 7,200 Mbps, Samsung's DDR5 memory chips can handle larger, more complex workloads with twice the performance of DDR4 chips and the ability to seamlessly handle 8K content.

"Samsung's DDR5 technology will enable the development of servers with twice as many cores, for extremely dense, cost- and energy-efficient implementation, with the memory bandwidth to support the performance and capacity that cloud native workloads demand," Rangarajan said.

"Organizations are grappling with how to handle increasingly vast amounts of data, and use it to provide value to customers without driving up power consumption and operational overhead," said In Dong Kim, Vice President of Memory Product Planning at Samsung. "Samsung and Ampere are committed to ongoing collaboration as we develop integrated cloud native solutions that power data-intensive services and applications today and into the future."

Learn more about Ampere's Cloud Native Processors and Samsung's market-leading memory solutions today.

SAMSUNG