

Transcript menu

Mitchell Haws	executive
Lisa Su	executive
Jean Hu	executive
Benjamin Reitzes	analyst
Aaron Rakers	analyst
Timothy Arcuri	analyst
Ross Seymore	analyst
Matthew Ramsay	analyst
Vivek Arya	analyst
Joseph Moore	analyst
Toshiya Hari	analyst
Stacy Rasgon	analyst
Harsh Kumar	analyst
Blayne Curtis	analyst
Thomas O'Malley	analyst
Christopher Danely	analyst

Operator

Greetings, and welcome to the AMD Second Quarter 2024 Conference Call. [Operator Instructions] As a reminder, this conference is being recorded. It is now my pleasure to introduce to you Mitch Haws, Vice President, Investor Relations. Thank you, Mitch.

You may begin.

Mitchell Haws

Thank you, and welcome to AMD's Second Quarter 2024 Financial Results Conference Call. By now, you should have had the opportunity to review a copy of our earnings press release and the accompanying slides.

If you have not had the chance to review these materials, they can be found on the Investor Relations page of amd.com.

We will refer primarily to non-GAAP financial measures during today's call. The full non-GAAP to GAAP reconciliations are available in today's press release and the slides posted on our website. Participants on today's conference call are Dr. Lisa Su, our Chair and Chief Executive Officer; and Jean Hu, our Executive Vice President, Chief Financial Officer, and Treasurer. This is a live call and will be replayed via webcast on our website.

Before we begin, I would like to note that Dr. Lisa Su will attend the Goldman Sachs Technology Communacopia and Technology Conference on Monday, September 9; and Mark Papermaster, Executive Vice President and Chief Technology Officer, will attend the Deutsche Bank Technology Conference on Wednesday, August 28. Today's discussion contains forward-looking statements based on current beliefs, assumptions and expectations, speak only as of today and as such, involve risks and uncertainties that could cause actual results to differ materially from our current expectations. Please refer to the cautionary statement in our press release for more information on factors that could cause actual results to differ materially. With that, I'll hand the call over to Lisa.

Lisa Su

Thank you, Mitch, and good afternoon to all those listening today. We delivered strong second quarter financial results, with revenue coming in above the midpoint of guidance and profitability increasing by a double-digit percentage driven by higher-than-expected sales of our Instinct, Ryzen, and EPYC processors.

We continued accelerating our AI traction as leading cloud and enterprise providers expanded availability of Instinct MI300X solutions, and we also saw positive demand signals for general purpose compute in both our client and server processor businesses.

As a result, second quarter revenue increased 9% year-over-year to \$5.8 billion as significantly higher sales of our Data Center and Client processors more than offset declines in Gaming and Embedded product sales.

We also expanded gross margin by more than 3 percentage points and grew EPS 19% as Data Center product sales accounted for nearly 50% of overall sales in the quarter.

Turning to the segments. Data Center segment revenue increased 115% year-over-year to a record \$2.8 billion, driven by the steep ramp of Instinct MI300 GPU shipments and a strong double-digit percentage increase in EPYC CPU sales. Cloud adoption remains strong as hyperscalers deploy 4th Gen EPYC CPUs to power more of their internal workloads and public instances.

We are seeing hyperscalers select EPYC processors to [indiscernible] portion of their applications and workloads, displacing incumbent offerings across their infrastructure with AMD solutions that offer clear performance and efficiency advantages. The number of AMD-powered cloud instances available from the largest providers has increased 34% from a year ago to more than 900.

We are seeing strong pull for these instances with both enterprise and cloud-first businesses.

As an example, Netflix and Uber both recently selected 4th Gen EPYC public cloud instances as one of the key solutions to power their mission-critical customer-facing workloads. In the enterprise, sell-through increased by a strong double-digit percentage sequentially. We closed multiple large wins in the quarter with financial services, technology, health care, retail, manufacturing, and transportation customers, including Adobe, Boeing, Industrial Light & Magic, Optiver, and Siemens.

Importantly, more than 1/3 of our enterprise server wins in the first half of the year were with businesses deploying EPYC in their data centers for the first time, highlighting our success attracting new customers while also continuing to expand our footprint with existing customers.

Looking ahead, our next-generation Turin family featuring our new Zen 5 core is looking very strong. Zen 5 is a grounds-up new core design optimized for leadership performance and efficiency. Turin will extend our TCO leadership by offering up to 192 cores and 384 threads, support for the latest memory and I/O technologies, and the ability to drop into existing 4th Gen EPYC platforms. We publicly previewed Turin for the first time in June, demonstrating our significant performance advantages in multiple compute-intensive workloads.

We also passed a major milestone in the second quarter as we started Turin production shipments to lead cloud customers. Production is ramping now ahead of launch, and we expect broad OEM and cloud availability later this year.

Turning to our Data Center AI business. We delivered our third straight quarter of record Data Center GPU revenue with MI300 quarterly revenue exceeding \$1 billion for the first time. Microsoft expanded their use of MI300X accelerators to power GPT-4 Turbo and multiple copilot services, including Microsoft 365 Chat, Word, and Teams. Microsoft also became the first large hyperscaler to announce general availability of public MI300X instances in the quarter. The new Azure VMs leverage the industry-leading compute performance and memory capacity of MI300X in conjunction with the latest ROCm software to deliver leadership inferencing price performance when running the latest frontier models, including GPT-4. Hugging Face was one of the first customers to adopt the new Azure instances, enabling enterprise and AI customers to deploy hundreds of thousands of models on MI300X GPUs with 1 click.

Our enterprise and cloud AI customer pipeline grew in the quarter, and we are working very closely with our system and cloud partners to ramp availability of MI300 solutions to address growing customer demand. Dell, HP, Lenovo, and Super Micro all have Instinct platforms in production, and multiple hyperscale and Tier 2 cloud providers are on track to launch MI300 instances this quarter.

On the AI software front, we made significant progress enhancing support and features across our software stack, making it easier to deploy high-performance AI solutions on our platforms.

We also continued to work with the open source community to enable customers to implement the latest AI algorithms.

As an example, AMD support for Flash Attention 2 algorithm was upstreamed, providing out-of-the-box support for AMD hardware in the popular library that can increase training and inference performance on large transformer models.

Our work with the model community also continued accelerating, highlighted by the launches of new models and frameworks with day 1 support for AMD hardware. At Computex, I was joined by the co-CEO of Stable Diffusion to announce that MI300 is the first GPU to support their latest SD 3.0 Image Generation LLM. Last week, we were proud to note that multiple partners use ROCm and MI300X to announce support for the latest Llama 3.1 models, including their 405 billion parameter version that is the industry's first frontier-level open source AI model. Llama 3.1 runs seamlessly on MI300 accelerators. And because of our leadership memory capacity, we're also able to run the FP16 version of the Llama 3.1 405B model in a single server, simplifying deployment and fine-tuning of the industry-leading model and providing significant TCO advantages. Earlier this month, we announced our agreement to acquire Silo AI, Europe's largest private AI lab with extensive experience developing tailored AI solutions for multiple enterprise and embedded customers, including Allianz, Ericsson, Finnair, Karber, Nokia, Philips, T-Mobile, and Unilever. The Silo team significantly expands our capability to service large enterprise customers looking to optimize their AI solutions for AMD hardware. Silo also brings deep expertise in large language model development, which will help accelerate optimization of AMD inference and training solutions.

In addition to our acquisitions of Silo AI, Mipsology, and Nod.ai, we have invested over \$125 million across a dozen AI companies in the last 12 months to expand the AMD AI ecosystem, support partners, and advance leadership AMD computing platforms.

Looking ahead, from a road map perspective, we are accelerating and expanding our Instinct road map to deliver an annual cadence of AI accelerators, starting with the launch of MI325X later this year. MI325X leverages the same infrastructure as MI300 and extends our generative AI performance leadership by offering twice the memory capacity and 1.3x more peak compute performance than competitive offerings. We plan to follow MI325X with the MI350 series in 2025 based on the new CDNA 4 architecture, which is on track to deliver a 35x increase in performance compared to CDNA 3. And our MI400 series powered by the CDNA Next architecture is making great progress in development and is scheduled to launch in 2026.

Turning to our AI solutions work. Broadcom, Cisco, HP Enterprise, Intel, Google, Meta, and Microsoft, all joined us to announce Ultra Accelerator Link, an industry standard technology to connect hundreds of AI accelerators that is based on AMD's proven Infinity Fabric technology. By combining UALink with the widely supported Ultra Ethernet consortium specification, the industry is coming together to establish a standardized approach for building the next

generation of high-performance data center AI solutions at scale. In summary, customer response to our multiyear Instinct and ROCm road maps is overwhelmingly positive and we're very pleased with the momentum we are building. We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner indicates agreement. See our [Cookie Policy](#) for more information.

As a result, we now expect Data Center GPU revenue to exceed \$4.5 billion in 2024, up from the \$4 billion we guided in April.

Turning to our Client segment. Revenue was \$1.5 billion, an increase of 49% year-over-year, driven by strong demand for our prior generation Ryzen processors and initial shipments of our next-generation Zen 5 processors. In PC applications, Zen 5 delivers an average of 16% more instructions per clock than our industry-leading previous generation of Ryzen processors.

For desktops, our upcoming Ryzen 9000 series processors dropped into existing AM5 motherboards and extends our performance and energy efficiency leadership across productivity, gaming, and content creation workloads.

For notebooks, we announced our Ryzen AI 300 series that extends our industry-leading CPU and GPU performance and introduces the industry's fastest NPU with [50 tops] of AI compute performance for Copilot+ PCs.

The first Ryzen AI 300 Series notebooks went on sale over the weekend to strong reviews, and more than 100 Ryzen AI 300 Series premium, gaming, and commercial platforms are on track to launch from Acer, ASUS, HP, Lenovo, and others over the coming quarters. Customer excitement for our new Ryzen processors is very strong, and we are well positioned for ongoing revenue share gains based on the strength of our leadership portfolio and design win momentum.

Now turning to our Gaming segment. Revenue declined 59% year-over-year to \$648 million as semi-custom SoC sales declined in line with our projections. Semi-custom demand remains soft as we are now in the fifth year of the console cycle, and we expect sales to be lower in the second half of the year compared to the first half. In gaming graphics, revenue increased year-over-year, driven by improved sales of our Radeon 6000 and 7000 Series GPUs in the channel.

Turning to our Embedded segment. Revenue decreased 41% year-over-year to \$861 million.

The first quarter marked the bottom for our Embedded segment revenue.

Although second quarter revenue was flattish sequentially, we saw early signs of order patterns improving and expect Embedded revenue to gradually recover in the second half of the year. Longer term, we are building strong design win momentum for our expanded Embedded portfolio. Design wins in the first half of the year increased by more than 40% from the prior year to greater than \$7 billion, including multiple triple-digit million-dollar wins, combining our Adaptive and x86 compute products. We announced our Alveo V80 accelerators that deliver leadership capabilities in memory-intensive workloads and entered early access on next-generation edge AI solutions with more than 30 key partners on our upcoming 2nd Gen Versal Adaptive SoCs. Last week, we also announced Victor Peng, President of AMD, would retire at the end of August. Victor has made significant contributions to Xilinx and AMD, including helping scale our Embedded business and leading our cross-company AI strategy. On a personal note, Victor has been a great partner to me, ensuring the success of our Xilinx acquisition and integration. On behalf of all of the AMD employees and Board, I want to thank Victor for all of his contributions to AMD's success and wish him all the best in his retirement. In summary, we delivered strong second quarter results and are well positioned to grow revenue significantly in the second half of the year, driven by our Data Center and Client segments.

Our Data Center GPU business is on a steep growth trajectory as shipments ramp across an expanding set of customers. We're also seeing strong demand for our next-generation Zen 5, EPYC, and Ryzen processors that deliver leadership performance and efficiency in both Data Center and Client workloads.

Looking ahead, the rapid advances in generative AI and development of more capable models are driving demand for more compute across all markets. Under this backdrop, we see strong growth opportunities over the coming years and are significantly increasing hardware, software, and solutions investments with a laser focus on delivering an annual cadence of leadership data center GPU hardware, integrating industry-leading AI capabilities across our entire product portfolio, enabling full stack software capabilities, amplifying our ROCm development with the scale and speed of the open source community, and providing customers with turnkey solutions that accelerate the time to market for AMD-based AI systems.

We are excited about the unprecedented opportunities in front of us and are well positioned to drive our next phase of significant growth.

Now I'd like to turn the call over to Jean to provide some additional color on our second quarter results. Jean?

Jean Hu

Thank you, Lisa, and good afternoon, everyone. I'll start with a review of our financial results and then provide our current outlook for the third quarter.

We are very pleased with our overall second quarter financial results that came in above expectations. On a year-over-year basis, Data Center segment revenue more than doubled, Client segment revenue grew significantly and we expanded the gross margin by 340 basis points.

For the second quarter of 2024, revenue was \$5.8 billion, up 9% year-over-year as revenue growth in the Data Center and the Client segments was partially offset by lower revenue in our Gaming and Embedded segment. Revenue increased 7% sequentially, primarily driven by growth in the Data Center and Client segments revenue. Gross margin was 53%, up 340 basis points year-over-year, primarily driven by higher Data Center revenue. Operating expenses were \$1.8 billion, an increase of 15% year-over-year as we continue to invest in R&D to address the significant AI growth opportunities ahead of us and enhanced go-to-market activities. Operating income was \$1.3 billion, representing a 22% operating margin. Taxes, interest expense and other was \$138 million. Diluted earnings per share was \$0.69, an increase of 19% year-over-year.

Now turning to our reportable segments. Starting with Data Center, Data Center delivered record quarterly segment revenue of \$2.8 billion, up 115%, a \$1.5 billion increase year-over-year. The Data Center segment accounted for nearly 50% of total revenue, led primarily by the steep ramp of AMD Instinct GPUs and a strong double-digit percentage EPYC server revenue growth. On a sequential basis, revenue increased 21%, driven primarily by strong momentum in AMD Instinct GPUs. Data Center segment operating income was \$743 million or 26% of revenue compared to \$147 million or 11% a year ago. Operating income was up more than 5x from the prior year, driven by higher revenue and operating leverage even as we significantly increased our investment in R&D. Client segment revenue was \$1.5 billion, up 49% year-over-year and 9% sequentially, driven primarily by AMD Ryzen processor sales. Client segment operating income was \$89 million or 6% of revenue compared to an operating loss of \$60 million a year ago. Gaming segment revenue was \$648 million, down 59% year-over-year and 20% sequentially. The decrease in revenue was primarily due to semi-custom inventory digestion and the

We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner indicates agreement. See our [Cookie Policy](#) for more information.

lower end market demand. Gaming segment operating income was \$77 million or 12% of revenue compared to \$225 million or 14% a year ago. Embedded segment revenue was \$861 million, down 41% year-over-year as customers continued to normalize their inventory levels. On a sequential basis, Embedded segment revenue was up 2%. Embedded segment operating income was \$345 million or 40% of revenue compared to \$757 million or 52% a year ago.

Turning to the balance sheet and cash flow.

During the quarter, we generated \$593 million in cash from operations, and free cash flow was \$439 million. Inventory increased sequentially by \$339 million to \$5 billion, primarily to support the continued ramp of Data Center GPU products. At the end of the quarter, cash, cash equivalents, and short-term investments were \$5.3 billion. In the second quarter, we returned \$352 million to shareholders, repurchasing 2.3 million shares, and we have \$5.2 billion of authorization remaining.

During the quarter, we retired \$750 million of debt that matured this past June, utilizing existing cash.

Now turning to our third quarter 2024 outlook.

We expect revenue to be approximately \$6.7 billion, plus or minus \$300 million. Sequentially, we expect revenue to grow approximately 15%, primarily driven by strong growth in the Data Center and Client segment.

We expect Embedded segment revenue to be up and the Gaming segment to decline by double-digit percentage. Year-over-year, we expect revenue to grow approximately 16%, driven by the steep ramp of our AMD Instinct processors, and a strong server and client revenue growth to more than offset the declines in the Gaming and the Embedded segment.

In addition, we expect third quarter non-GAAP gross margin to be approximately 53.5%, non-GAAP operating expenses to be approximately \$1.9 billion, non-GAAP effective tax rate to be 13%, and the diluted share count is expected to be approximately 1.64 billion shares. Also during the third quarter, we expect to close the acquisition of Silo AI for approximately \$665 million in cash. In closing, we made significant progress during the quarter toward achieving our financial goals. We delivered record MI300 revenue that exceeded \$1 billion and demonstrated solid traction with our next-gen Ryzen and EPYC product. We expanded gross margin significantly and drove earnings growth while increasing investment in AI.

Looking forward, the opportunity [indiscernible] will remain focused on executing to our long-term growth strategy while driving financial discipline and operational excellence. With that, I'll turn it back to Mitch for the Q&A session.

Mitchell Haws

Thank you, Jane. John, we're happy to poll the audience for questions.

Operator

[Operator Instructions] And the first question comes from the line of Ben Reitzes with Melius Research.

Benjamin Reitzes

Congratulations on these results. Lisa, I wanted to ask you about MI300, how you see it playing out sequentially for the rest of the year. I guess there's about \$2.8 billion left to hit your annual target.

So I'm wondering if you see things picking up in the fourth quarter and how that's going sequentially. And if you don't mind, I wanted to also ask about next year if you see potential for rapid growth.

You're probably aware of some of the chatter out there, and I just was wondering if you're already seeing signs that you can grow significantly, given your road map for next year.

Lisa Su

Yes. Great, Ben. Thanks for the question.

So first of all, on sort of MI300 and the customer evolution, we're very happy with how MI300 has progressed. When we started the year, I think the key point for us was to get our products into our customers' data centers, to have them qualify their workloads, to really ramp in production and then see what the production capabilities are, especially performance and all of those things. And I can say now being sort of more than halfway through the year, we've seen great progress across the board.

As we look into the second half of the year, I think we would expect that MI300 revenue would continue to ramp in the third quarter and the fourth quarter. And we're continuing to expand both current deployments with our existing customers as well as we have a large pipeline of customers that we're working through that are getting familiar with our architecture and software and all that stuff.

So I would say overall, very pleased with the progress, and really continuing right on track to what we expected from the capabilities of the product.

As we go into next year, I mean, one of the important things that we announced at Computex was increasing and expanding our road map.

I think we feel really good about our road map. We're on track to launch MI325 later this year, and then next year, our MI350 Series, which will be very competitive with Blackwell solutions. And then we're well on our way to our CDNA Next as well.

So I think, overall, we remain quite bullish on the overall AI market.

I think the market continues to need more compute. And we also feel very good that our hardware and software solutions are getting good traction, and we're continuing to expand that pipeline.

And the next question comes from the line of Aaron Rakers with Wells Fargo.

Aaron Rakers

Congrats on the quarter as well. I guess sticking on the Data Center side, as we look forward and you think about the full year, I'm curious of how you're currently thinking about the EPYC server CPU growth expectations as we go forward. And any kind of updated thoughts on your ability to kind of continue to gain share in the server market? Just kind of just update us on how you see the server market playing out over the next couple of quarters.

Lisa Su

Yes, sure, Aaron. Thanks for the question.

So we're very pleased with the progress that we've made with EPYC.

I think a couple of things.

First of all, in terms of competitive positioning and just the traction in the market, our 4th Gen EPYC between Gen 1 Bergamo is really doing very well. We've seen broad adoption across cloud. And then we've been very focused on enterprise as well as third-party cloud instances. And as I said in the prepared remarks, we're starting to see very nice traction in enterprise with both new customers as well as existing customers, and then for third-party cloud adoption, also a good pickup there as well.

So I think overall, I think our EPYC portfolio has done well.

Going into the second half of the year, I think we also feel good about it. There are a couple of positives there. We see -- first of all, the market looks like it's improving, so we have seen some return to spending in both enterprise and cloud.

And so I think those are positive market trends. And then in addition to that, we are in the process of launching Turin.

So we started production here in the second quarter and we're on track to launch broadly in the second half of the year. We'll see some revenue of Turin in the second half of the year contributing as well.

So overall, I think the server market and our ability to continue to grow share in the server market is one of the things that we see in the second half of the year.

Operator

And the next question comes from the line of Timothy Arcuri with UBS.

Timothy Arcuri

Lisa, I wanted to ask about the Data Center GPU road map.

As you said, 325 launching later this year, so I guess I had 2 questions. Does the greater than \$4.5 billion, does that include any revenue from 325? And can you talk a little bit more about 350? Obviously, we're seeing a big rack scale or shift toward rack-scale systems for the competition's product. And I'm wondering if that's what 350 is going to look like. Is it going to have liquid cooling and is it going to have a rack scale aspect to it?

Lisa Su

Yes, absolutely.

So let me start with your original question. I mean, I think looking at 325X, we are on track to launch later this year. From a revenue standpoint, there will be a small contribution in the fourth quarter but it really is still mostly the MI300 capabilities. And 325 will start in the fourth quarter and then ramp more in the first half of next year. And then as we look at the 350 Series, what we're seeing and the reason we call it a series is because there will be multiple SKUs in that series that we'll go through the range of, let's call it, air-cooled to liquid-cooled. In spending time with our customers.

I think there are people who certainly want more rack level solutions, and we're certainly doing much more in terms of system-level integration for our products.

You'll see us invest more in system-level integration. But we also have many customers who want to use their current infrastructure.

I think the beauty of the MI350 series is it actually fits into the same infrastructure as the MI300 series.

And so it would lend itself to, let's call it, a pretty fast ramp if you've already invested in 300 or 325.

So we see the range of options, and that's part of the expansion of the road map that we're planning.

Operator

And the next question comes from the line of Ross Seymore with Deutsche Bank.

Ross Seymore

Congrats on the strong results. Well, Data Center is obviously very important. I just want to pivot to the Client side. Lisa, can you talk about the AI PC side of things? How you believe AMD is positioned? Are you seeing any competitive intensity changing with the emergence of ARM-based systems? Just wanted to see how you're expecting that to roll out and what it means to second half seasonality.

Lisa Su

Yes. Sure, Ross.

We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner So indicates agreement. See our [Cookie Policy](#) for more information.

I think we have a very strong road map, so I'm very pleased with the road map. The Zen 5-based products, we're launching both notebook and desktop in this -- in the middle of this year. What we've seen is actually very positive feedback on the product.

So we just actually launched the [first strict spaced] notebooks over the weekend. They went on sale.

You may have seen some of the reviews. The reviews are very positive.

Our view of this is the AI PC is an important add to the overall PC category.

As we go into the second half of the year, I think we have better seasonality in general, and we think we can do, let's call it, above-typical seasonality, given the strength of our product launches and when we're launching. And then into 2025, you're going to see AI PCs across sort of a larger set of price points, which will also open up more opportunities.

So overall, I would say the PC market is a good revenue growth opportunity for us. The business is performing well. The products are strong. And we're working very closely with both the ecosystem partners as well as our OEM partners to have strong launches here into the second half of the year.

Ross Seymore

And is the ARM side changing anything or not really?

Lisa Su

Look, I think at this point, the PC market is a big market and we are underrepresented in the market. I would say that we take all of our competition very seriously. That being the case, I think our products are very well positioned.

Operator

And the next question comes from the line of Matt Ramsay with Cowen.

Matthew Ramsay

Lisa, I wanted to maybe draw a parallel between the Instinct portfolio that your company is rolling out now and what you guys did 5 or 6 years ago with EPYC. And I remember when the Naples product launched, there was a lot of, I would say, reaction positively and negatively and sort of sentiment around where your road map might go to relatively small perturbations in what the volumes were. Super early, but if I remember back to that, what was the most important was that was the toehold into the market for long-term engagement, both on the software side and the hardware side with your customers 2, 3, 4 generations forward.

So is that an accurate parallel to where you guys are with MI300? And maybe you could talk about the level of engagement, the intensity of engagement, the breadth of it across the customer base with 350 and 400.

Lisa Su

Yes, absolutely, Matt.

So look, as I said earlier, we're very pleased with the progress that we're making on the Instinct road map. This is absolutely a long-term play so absolutely, you're correct. It has a lot of parallels to the EPYC journey, where you really have to -- you gain more opportunities, broader workloads, larger deployments as you go from generation to generation.

So we are playing the long game here.

Our conversations with our customers, so I would start with first, in the near term, we had some very key milestones that we wanted to pass this year. And as I said, they related to getting hardware in volume in multiple hyperscalers as well as large Tier 2 customers. We've done that. We've now seen our software in a lot of different environments, and it's matured substantially. ROCm is in very, from a standpoint of features, functions, out-of-box performance, getting to performance with customers, we've gained a lot of confidence and learned a lot in that whole process. The networking aspects of building out the rack scale and the system-level items are areas that we're continuing to invest in. And then the point of having long-term conversations across multiple generations is also really important.

So I think all of those things have progressed well. We view this as very good progress for MI300, but we have a lot more to do. And I think the various road maps will help us open up those opportunities over the next couple of years.

Operator

And the next question comes from the line of Vivek Arya with Bank of America Securities.

Vivek Arya

Lisa, there seems to be this ongoing industry debate about AI monetization and whether your customers are getting the right ROI on their CapEx. And today, they have these 3 options, right? They can buy GPUs from your largest competitor with all the software bells and whistles and incumbency, or they can do custom chips or they can buy from AMD.

So how do you think this plays out next year? Do you think your customers, given all this concern around monetization, does it make them consolidate their CapEx around just the other 2 suppliers? How is your visibility going into next year, given this industry debate? And how will AMD continue to kind of carve a position between these 2 other competitive choices that are out there?

Lisa Su

Yes, sure, Vivek. Well, I mean, I think you talk to a lot of the same people that we talk to.

So I think the investment cycle will continue to be strong. And then relative to the various choices for the size of the market, I firmly believe that there will be multiple solutions, whether you're talking about GPUs or you're talking about custom chips or ASICs, there will be multiple solutions. In our case, I think we've demonstrated a really strong road map and the ability to partner well with our customers. And from the standpoint of that deep engagement, hardware, software co-optimization is so important in that. And for large language models, GPUs are still the architecture of choice.

So I think the opportunity is very large. And I think our piece of that is really strong technology with strong partnerships with the key AI market makers.
Operator

And the next question comes from the line of Joe Moore with Morgan Stanley.

Joseph Moore

I also wanted to ask about MI300. I wonder if you could talk about training versus inference. Do you have a sense -- I know that a lot of the initial focus was inference, but do you have traction on the training side? And any sense of what that split may look like over time?

Lisa Su

Yes, sure. Thanks for the question, Joe.

So as we said on MI300, there are lots of great characteristics about it.

One of it -- one of them is our memory bandwidth and memory capacity is leading the industry. From that standpoint, the early deployments have largely been inference in most cases, and we've seen fantastic performance from an inference standpoint.

We also have customers that are doing training. We've also seen that from a training standpoint, we've optimized quite a bit our ROCm software stack to make it easier for people to train on AMD. And I do expect that we'll continue to ramp training over time.

As we go forward, I think you'll see -- the belief is that inference will be larger than training from a market standpoint. But from an AMD standpoint, I would expect both inference and training to be growth opportunities for us.

Operator

And the next question comes from the line of Toshiya Hari with Goldman Sachs.

Toshiya Hari

I had a question on the MI300 as well. Curiously, if you're currently shipping to demand or if the updated annual forecast of \$4.5 billion is in some shape or form supply constrained, think last quarter, you gave some comments on HBM and CoAs. Curious if you could provide an update there. And then my part B to my question is on profitability for MI300.

I think in the past, you've talked about the business being accretive and improving further over time as you sort of work through the kinks, if you will. Has that view evolved or changed at all, given sort of the competitive intensity and your need to invest, whether it be through organic R&D or some of the acquisitions you've made? Or are you still confident that profit margins in the business continue to expand?

Lisa Su

Yes. Sure, Toshiya. Thanks for the question.

So on the supply side, let me make a couple of comments and then maybe I'll let Jean comment on sort of the trajectory for the business.

So on the supply side, we made great progress in the second quarter. We ramped up supply significantly exceeding \$1 billion in the quarter.

I think the team has executed really well.

We continue to see line of sight to continue increasing supply as we go through the second half of the year. But I will say that the overall supply chain is tight and will remain tight through 2025.

So under that backdrop, we have great partnerships across the supply chain. We've been building additional capacity and capability there.

And so we expect to continue to ramp as we go through the year. And we'll continue to work both supply as well as demand opportunities, and really that's accelerating our customer adoption overall, and we'll see how things play out as we go into the second half of this year.

Jean Hu

Yes. On your second question about the profitability, first, our team has done a tremendous job to ramp the product MI300. It's a very complex product so we ramped it successfully. At the same time, the team also started to implement operational optimization to continue to improve gross margin.

So we continue to see the gross margin improvement. Over time, in the longer term, we do believe gross margin will be accretive to corporate average. From a profitability perspective, AMD always invests in platforms.

If you look at our Data Center platform, especially both the server and the Data Center GPU side, we are ramping the revenue. The business model can leverage very significantly even from GPU side. Because the revenue ramp has been quite significant, the operating margin continued to expand. We definitely want to continue to invest as the opportunity is huge. At the same time, it is a profitable business already.

Operator

And the next question comes from the line of Stacy Rasgon with Bernstein Research

So with Gaming down double digits, it probably means you've got close to \$1 billion of growth revenue across Data Center, Client, and Embedded. I was wondering if you could give us some color on how that \$1 billion-ish splits out across those 3 businesses. Like if I had 70% of it going to Data Center and 20% going to Client and 10% going to Embedded, like would that be like way off? Or how should we think about that apportioning out across the segment?

Lisa Su

Yes. Maybe Stacy, let me give you the following color.

So the Gaming business is down double digit as you state. Think of it as the Data Center is the largest piece of it, Client next. And then on the Embedded side, think of it as single-digit sequential growth.

Stacy Rasgon

Got it.

So I mean within that Data Center piece then, how does that split out? I mean, is the bulk of it [indiscernible] Instinct? Or is it sort of equally weighted between Instinct and EPYC? Or like again, how does it -- again, if you got, I don't know, \$400 million to \$600 million of sequential Data Center growth or something like that, how does it split up?

Lisa Su

Yes.

So again, without being that granular, we will see both -- certainly, the Instinct GPUs will grow and we'll see also very nice growth on the server side.

Operator

And the next question comes from the line of Harsh Kumar with Piper Sandler.

Harsh Kumar

Lisa, from my rudimentary understanding, the large difference between your Instinct products and the adoption versus your nearest competitor is kind of rack level performance and that rack level is the structure that you may be lacking.

You talked a little bit about UALink. I was wondering if you could expand on that and give us some more color on when that might -- when that gap might be closed. Or is this a major step for the industry to close that gap? Just any color would be appreciated.

Lisa Su

Yes.

So Harsh, overall, maybe if I take a step back and just talk about how the systems are evolving, there's no question that the systems are getting more complex, especially as you go into large training clusters, and our customers need help to put those together. And that includes the sort of Infinity Fabric-type solutions that are the basis for the UALink things as well as just general rack level system integration.

I think what you should expect, Harsh is, first of all, we're very pleased with all of the partners that have come together for UALink. We think that's an important capability. But we have all of the pieces of this already within sort of the AMD umbrella with our Infinity Fabric, with the work with our networking capability through the acquisition of Pensando. And then you'll see us invest more in this area.

So this is part of how we help customers get to market faster is by investing in all of the components, so the CPUs, the GPUs, the networking capability as well as system-level solutions.

Operator

And the next question comes from the line of Blayne Curtis with Jefferies.

Blayne Curtis

I just want to ask another question on MI300.

Just curious if you can kind of characterize the makeup of the customers in the first half. I know you had, end of last year, a government customer. Is there still a government contingency? And kind of the second part of it is really you've invested in all these software assets. Kind of curious the challenge of ramping the next wave of customers. I know there's been a lot of talk on some hardware challenges, memory issues and such, but then you're investing in software. I'm sure that's a big challenge, too.

Just kind of curious what the biggest hurdle is for you to kind of get that next wave of customers ramp.

Lisa Su

Yes.

So Blayne, a lot of pieces to that question so let me try to address them.

First, on your question about, I think you're basically asking about the supercomputing piece. That was mainly Q4 and a bit in Q1.

So if you think about our Q2 revenue, think about it as almost all AI.

So it's MI300X, it's for large AI, hyperscalers as well as OEM customers going to enterprise and Tier 2 data centers.

So that's the makeup of the customer set. And then in terms of the various pieces of what we're doing, I think first on your question about memory, I think there's a lot of noise in the system. I wouldn't really pay attention to all that noise in the system. I mean, this has been an incredible ramp. And I'm actually really proud of what the team has done in terms of just definitely fastest product ramp that we've ever done to \$1 billion here in the -- over \$1 billion in the second quarter and then ramping each quarter in Q3 and Q4.

In terms of memory, we have multiple suppliers that we've qualified on HBM3. And it's a tricky -- memory is a tricky business but I think we've done it very well and that's there. And then we're also qualifying HBM3E for future products with multiple memory suppliers as well.

So to your overarching question of what are the things that we're doing, the exciting part of this is that the ROCm capability has really gotten substantially better because so many customers have been using it. And with that, what we look at is out-of-box performance, how long does it take a customer to get up and running on MI300? And we've seen, depending on the software that companies are using, particularly if you're based on some of the higher-level frameworks like PyTorch, et cetera, we can be out-of-the-box running very well in a very short amount of time, like, let's call it, very small number of weeks. And that's great because that's expanding the overall portfolio.

We are going to continue to invest in software, and that was one of the reasons that we did the Silo AI acquisition. It's a great acquisition for us. 300 scientists and engineers. These are engineers that have experience with AMD hardware and are very, very good at helping customers get up and running on AMD hardware. And that's -- so we view this as the opportunity to expand the customer base with talent like Silo AI, like Nod.ai which brought a lot of compiler talent. And then we continue to hire quite a bit organically.

So I think Jean said earlier that we see leverage in the model, but we're going to continue to invest because this opportunity is huge, and we have all of the pieces. This is just about building out scale.

Operator

And the next question comes from the line of Tom O'Malley with Barclays.

Thomas O'Malley

I'll give you a breather from the MI300 for a second, but just to focus on Client in the second half. No problem. Focused on Client in the second half, you kind of said above-seasonal for September, December.

You're obviously launching a new notebook, desktop product, but you're also talking about AI PC. Could you just break down where you're seeing those above-seasonal trends? Is it the ASP uplift you're getting from the new products? Is it a unit assumption that's coming with AI PC? Just any kind of breakdown between those 2 and why you're seeing it a little bit better.

Lisa Su

Sure, Tom.

So I think you actually said it well.

We are launching Zen 5 desktops and notebooks with volume ramping in the third quarter. And that's the primary reason that we see above-seasonal. The AI PC element is certainly 1 element of that, but there is just the overall refresh. Usually, desktop launches going into a third quarter are good for us, and we feel that the products are very well positioned.

So those are the primary reasons.

Operator

And our final question comes from the line of Chris Danely with Citi.

Christopher Danely

Just a question on gross margin.

So if we look at your guidance, it seems like the incremental gross margin is dropping a little bit for Q3. Why is that happening? And then just a follow-up on another part of the gross margin angle. Have you changed your gross margin expectations for the MI300? Has the accretion point moved out a little bit?

Jean Hu

Yes, Chris, thanks for the question.

I think, first, we have made a lot of progress, as you mentioned, this year to expand our gross margin from 2023 at a 50 percentage point to, we actually guided 53.5% for Q3. The primary driver is really the faster Data Center business growth.

If you look at the Data Center business as a percentage of revenue from 37% in Q4 last year to now close to 50%. That faster expansion really helped us with the gross margin. When you look at the second half, we'll continue to see Data Center to be the major driver of our top line revenue growth, will help with the margin expansion. But there are some other puts and takes.

I think Lisa mentioned the PC business actually is going to do better in second half, especially typically, seasonally, it tends to be more consumer-focused.

So that really is a little bit different dynamics there.

Secondly, I would say Embedded business, we are going to see Embedded business to be up sequentially each quarter. But the recovery, as we mentioned earlier, is more gradual.

We use cookies on this site to provide a more responsive and personalized service. Continuing to browse, clicking I Agree, or closing this banner indicates agreement. See our [Cookie Policy](#) for more information.

As far as MI300, we are quite confident over the long term, it will be accretive to our corporate average. We feel pretty good about the overall Data Center business to continue to be absolutely the driver of gross margin expansion.

Operator

I would like to turn the floor back over to Mitch for any closing comments.

Mitchell Haws

Great. That concludes today's call. Thanks to all of you for joining us today.

Operator

And ladies and gentlemen, that does conclude today's teleconference.

You may disconnect your lines at this time. Thank you for your participation.