Amazon.com, Inc. (NASDAQ:AMZN) Q3 2023 Earnings Conference Call October 26, 2023 5:30 PM ET

Company Participants

Dave Fildes - Director of Investor Relations
Andrew Jassy - President & Chief Executive Officer
Brian Olsavsky - Senior Vice President & Chief Financial Officer

Conference Call Participants

Justin Post - Bank of America
Doug Anmuth - JPMorgan Chase & Co.
Brian Nowak - Morgan Stanley
Eric Sheridan - Goldman Sachs Group
Mark Mahaney - Evercore ISI

Operator

Thank you for standing by. Good day everyone and welcome to the Amazon.com Third Quarter 2023 Financial Results Teleconference. [Operator Instructions] Today's call is being recorded.

And for opening remarks, I will be turning the call over to the Vice President of Investor Relations, Dave Fildes. Thank you, sir. Please go ahead.

Dave Fildes

Hello and welcome to our Q3 2023 financial results conference call. Joining us today to answer your questions is Andy Jassy, our CEO; and Brian Olsavsky, our CFO. As you listen to today's conference call, we encourage you to have our press release in front of you which includes our financial results as well as metrics and commentary on the quarter. Please note, unless otherwise stated, all comparisons in this call will be against our results for the comparable period of 2022. Our comments and responses to your questions reflect management's views as of today, October 26, 2023, only and will include forward-looking statements. Actual results may differ materially. Additional information about factors that could potentially impact our financial results is included in today's press release and our filings with the SEC, including our most recent annual report on Form 10-K and subsequent filings.

During this call, we may discuss certain non-GAAP financial measures in our press release, slides accompanying this webcast and our filings with the SEC, each of which is posted on our IR website. You will find additional disclosures regarding these non-GAAP measures, including reconciliations of these measures with comparable GAAP measures. Our guidance incorporates the order trends that we've seen to date and what we believe today to be appropriate assumptions. Our results are inherently unpredictable and may be materially affected by many factors, including fluctuations in foreign exchange rates, changes in global economic and geopolitical conditions and customer demand and spending, including the impact of

recessionary fears, inflation, interest rates, regional labor market constraints, world events, the rate of growth of the Internet, online commerce, cloud services and new and emerging technologies and the various factors detailed in our filings with the SEC. Our guidance assumes, among other things, that we don't conclude any additional business acquisitions, restructurings or legal settlements. It's not possible to accurately predict demand for our goods and services and therefore, our actual results could differ materially from our guidance.

And now, I'll turn the call over to Andy.

Andrew Jassy

Thanks, Dave. Today, we're reporting $143.1 billion in revenue, up 11% year-over-year; $11.2 billion in operating income, up 343% year-over-year or $8.7 billion; and $20.2 billion in trailing 12-month free cash flow adjusted for equipment finance leases which is up $41.7 billion versus the comparable period last year. We continue to be encouraged by the progress we're making in lowering our cost to serve, improving our customer experiences and investing for future growth.

I'll start with our stores business. Our move earlier this year from a single national fulfillment network in the U.S. to 8 distinct regions represented one of the most significant changes to our fulfillment network in our history. This change has gone more smoothly and made more impact than we optimistically expected. And you can see the benefits in many forms. Regional fulfillment clusters with higher local in-stock levels and optimized connections between fulfillment centers and delivery stations mean shorter distances and fewer touches to get items to customers.

Shorter travel distances and fewer touches mean lower cost to serve. But perhaps most importantly, shorter distances and fewer touches mean that customers are getting their shipments faster. We remain on pace to deliver the fastest delivery speeds for Prime customers in our 29-year history. And as I talked about last quarter, we know how important speed of delivery is to customer satisfaction and buying behavior. A good example is the significant growth we're seeing in consumables and everyday essentials. When customers are getting items as quickly and conveniently as they are now from Amazon, they're going to consider us more frequently for more of their shopping needs. As we've shared the last few quarters, we've re-evaluated every part of our fulfillment network over the last year.

The first substantial rearchitecture centered on the regionalization change. We obviously like the results but don't think we fully realize all the benefits yet and we continue to make steady improvements in fine-tuning the placement algorithms to enable even more in-region fulfillment and to further increase consolidation into fewer shipments. We've also identified several substantial changes to our inbound processes that we believe could have a significant impact on our cost to serve and speed of delivery. We have a long way before being out of ideas to improve cost and speed. The team is really humming on this and I'm proud of the way they're inventing and executing together.

Moving to AWS and our investments in generative AI. AWS revenue grew 12% year-over-year in Q3, with $919 million of incremental quarter-over-quarter revenue and now has the annualized revenue run rate of $92 billion. AWS' year-over-year growth rate continued to stabilize in Q3. And while we still saw elevated cost optimization relative to a year ago, it's continued to attenuate as more companies transition to deploying net new workloads.

Companies have moved more slowly in an uncertain economy in 2023 to complete deals. But we're seeing the pace and volume of closed deals pick up and we're encouraged by the strong last couple of months of new deals signed. For perspective, we signed several new deals in September with an effective date in October that won't show up in any GAAP reported number for Q3 but the collection of which is higher than our total reported deal volume for all of Q3. Deal signings are always lumpy and the revenue happens over several years but we like the recent deal momentum we're seeing.

Top of mind for most companies continues to be generative AI. As I mentioned last quarter, we think about generative AI as having 3 macro layers, each of which is very large in each of which we're investing. A few updates there. At the lowest layer is the compute to train large language models, or LLMs and produce inferences or predictions. The key to this compute is the chip inside it. As we've shared, we've been working on custom silicon for training and inference with our Trainium and Inferentia chips, respectively.

Recently, we announced that leading LLM maker Anthropic chose AWS as its primary cloud provider and will use Trainium and Inferentia to build, train and deploy its future LLMs. As part of this partnership, AWS and Anthropic will collaborate on the future development of Trainium and Inferentia technology. We believe this collaboration will be helpful in continuing to accelerate the price performance advantages that Trainium and Inferentia deliver for customers.

In the middle layer which we think of as large language models as a service, we recently introduced general availability for Amazon Bedrock which offers customers access to leading LLMs from third-party providers like Anthropic, Stability AI, coherent AI 21 as well as from Amazon's own LLMs called Titan, where customers can take those models, customize them using their own data but without leaking that data back into the generalized LLM, have access to the same security, access control and features that they run the rest of their applications within AWS all through a managed service.

In the last couple of months, we've announced the imminent addition of Meta's Llama 2 model to Bedrock, the first time it's being made available through a fully managed service. Also through our expanded collaboration with Anthropic, customers will gain access to future Anthropic models through Bedrock with exclusive early access to unique features, model customization and the ability to fine-tune the models. And Bedrock has added several new compelling features, including the ability to create agents which can be programmed to accomplish tasks like answering questions or automating workflows.

In these early days of generative AI, companies are still learning which models they want to use which models they use for what purposes and which model sizes they should use to get the

latency and cost characteristics they desire. In our opinion, the only certainty is that there will continue to be a high rate of change. Bedrock helps customers with this fluidity, allowing them to rapidly experiment with move between model types and sizes and enabling them to pick the right tool for the right job. The customer reaction to Bedrock has been very positive and the general availability is buoyed that further. Bedrock is the easiest way to build and scale enterprise-ready generative AI applications and a real game changer for developers and companies trying to get value out of this new technology.

And the top layer which are the applications that run the LLMs, our generative AI coding companion Amazon CodeWhisperer has gotten a lot of early traction and got a lot more powerful recently with the launch of its new customization capability. The number one enterprise request for coding companions has been wanting these companions to be familiar with customers' proprietary code bases. It's not just having code companions trained in open source code, companies want the equivalent of a long-time senior engineer who knows their code base well. That's what CodeWhisperer just launched, another first of its kind out there in its current form and customers are excited about it.

A few last comments on AWS's generative AI work. As you can tell, we're focused on doing what we've always done for customers, taking technology that can transform customer experiences and businesses but they can be complex and expensive and democratizing it for customers of all sizes and technical abilities. It's also worth remembering that customers want to bring the models to their data, not the other way around. And much of that data resides in AWS as the clear market segment leader in cloud infrastructure. We're innovating and delivering at a rapid rate and our approach is resonating with customers.

The number of companies building generative AI apps in AWS is substantial and growing very quickly, including Adidas, Booking.com, Bridgewater, Clariant, GoDaddy, LexisNexis, Merck, Royal Philips and United Airlines, to name a few. We are also seeing success with generative AI start-ups like Perplexity.ai who chose to go all in with AWS, including running future models in Trainium and Inferentia. And the AWS team has a lot of new capabilities to share with its customers at its upcoming AWS re:Invent conference.

Beyond AWS, all of our significant businesses are working on generative AI applications to transform their customer experiences. There are too many for me to name on this call but a few examples include, in our stores business, we're using generative AI to help people better discover products they want to more easily access the information needed to make decisions. We use generative AI models to forecast inventory we need in our various locations and to derive optimal last mile transportation routes for drivers to employ.

We're also making it much easier for our third-party sellers to create new product pages by entering much less information and letting the models to the rest. In advertising, we just launched a generative AI image generation tool, where all brands need to do is upload a product photo and description to quickly create unique lifestyle images that will help customers discover products they love. And in Alexa, we built a much more expansive LLM and previewed the early version of this.

Apart from being a more intelligent version of herself, Alexa's new conversational AI capabilities include the ability to make multiple requests at once as well as more natural and conversational requests without having to use specific phrases. We continue to be convicted that the vision of being the world's best personal assistant is a compelling and viable one and that Alexa has a good chance to be one of the long-term winners in this arena. Every one of our businesses is building generative AI applications to change what's possible for customers and we have a lot more to come.

We're also encouraged by the progress we're making in our newer initiatives. Just to name a few, we're pleased with what we're seeing in Prime Video. Prime Video continues to be an integral part of the Prime value proposition where it's often one of the top 2 drivers of customers signing up for Prime. We also have increasing conviction that Prime Video can be a large and profitable business in its own right as we continue to invest in compelling exclusive content for prime members but also offer the best selection of premium streaming video content anywhere with our marketplace offering, including channels for customers who can subscribe to channels like Max, Paramount+, BET Plus and MGM+, as well as our broad transaction video-on-demand selection.

As we continue to invest in compelling content, beginning in early 2024, Prime Video shows and movies will include limited advertisements. We aim to have meaningfully fewer ads than linear TV and other streaming TV providers. If customers prefer and add free option, we plan to offer that for an additional $2.99 per month for U.S. members. There is still a lot of work to be done in innovation ahead but we're excited about our future on Prime Video.

We're seeing progress on a number of our investments that expand our ability to serve more consumers and sellers in their e-commerce missions. Our emerging international stores continue to improve their customer experiences and profitability and are on a strong trajectory. Both consumers and sellers are excited about Buy with Prime which enables third-party sellers with direct-to-consumer websites to offer Amazon Prime members the same fast payments and delivery options they receive on Amazon.com.

We recently announced the capability for sellers to integrate Buy with Prime with their Shopify account, making it easier for Shopify merchants to manage their businesses with inventory pricing and promotions automatically synced in one place. And we're seeing very positive early response from sellers to Supply Chain by Amazon, a fully automated set of supply chain services where Amazon can pick up inventory from manufacturing facilities around the world, ship it across borders, handle customs clearance and ground transportation, store inventory in bulk, manage replenishment across Amazon and other sales channels and deliver directly to customers, all without sellers having to worry about managing their supply chain.

Our health care team is continuing to make health care easier for people to access. The Amazon Pharmacy customer experience has significantly evolved this year and customers are responding that both in their purchasing behavior and qualitative feedback. We built RXPass for customers to get unlimited supply of eligible medications for $5 per month, meaningfully reduce the cost for customers to get insulin and diabetes products and partnered with Blue Shield of

California to offer a first-of-its-kind model to provide more affordable pharmacy care to its 4.8 million members, providing fast and free delivery of prescription medications and 24/7 access to pharmacists. We remain convinced that we can be part of the solution of making health care a better customer experience.

And our low Earth orbit satellite initiative Project Kuiper which aims to bring fast, affordable broadband to underserved communities around the world, took a meaningful step forward in the last few weeks with the successful launch of 2 prototype satellites. We will use this multi-month mission to test our satellites and network from space and collect data ahead of the planned start of satellite production later this year.

I'd like to close by thanking our teams around the world who are gearing up for 2 of our most significant events across the company. First, our annual AWS re:Invent conference that begins on November 27. The team is excited to share a lot of new capabilities with customers and provide an array of opportunities for builders to learn and connect with one another. And on the stores side, we've already kicked off what will be our 29th holiday shopping season. Prime Big Deal Days held earlier this month was our most successful October holiday kick-off event ever, with Prime members saving more than $1 billion across hundreds and millions of items sold. Just as we do all year long, we aim to make our customers' lives easier and better every day and there's no time where it's more important to us that we deliver on this mission than during the busy holiday shopping season.

With that, I'll turn it over to Brian.

Brian Olsavsky

Thanks, Andy. Overall, we saw a strong performance in the third quarter. Worldwide revenue was $143.1 billion, representing an increase of 11% year-over-year, excluding the impact of foreign exchange and approximately $100 million above the top end of our guidance range. We saw our highest quarterly worldwide operating income ever which was $11.2 billion for the quarter, an increase of $8.7 billion year-over-year from $2.7 billion above the high end of our guidance range.

North America revenue was $87.9 billion, an increase of 11% year-over-year. And international revenue was $32.1 billion, an increase of 11% year-over-year, excluding foreign exchange. During the quarter, we held our biggest Prime Day event ever with prime members purchasing more than 375 million items worldwide and saving more than $2.5 billion on millions of deals across the Amazon store. Outside of Prime Day, we continue to see strong demand across everyday essentials, including categories like beauty and health and personal care.

From a customer behavior standpoint, we still see customers remaining cautious about price, trading down where they can and seeking out deals, coupled with lower spending on discretionary items. Building on the momentum from last quarter, we set another record for delivery speed. For the year-to-date period through the third quarter, we have delivered at the

fastest speeds ever in the United States. These improvements in delivery speeds have been a key driver of growth and are resulting in increased purchase frequency by our Prime members.

Third-party sellers grew at 18% year-over-year, excluding foreign exchange, primarily driven by selection expansion and growing adoption of our optional services for sellers, including Fulfillment by Amazon and paid account management and more. During the quarter, we hosted Amazon Accelerate, our annual seller conference, where we launched a number of new innovations and product developments for our sellers, including Supply Chain by Amazon.

We also continue to see durable growth in advertising which grew 25% year-over-year, excluding foreign exchange, primarily driven by sponsored products as we lean into machine learning to improve the relevancy of the ads we show our customers and enhance our measurement capabilities on behalf of advertisers. We have seen strong improvement in our profitability. North America operating income was $4.3 billion, an increase of $4.7 billion year-over-year, resulting in an operating margin of 4.9%, up 100 basis points quarter-over-quarter. Since North America operating margins bottomed out in Q1 of 2022, we have now seen 6 consecutive quarters of improvement, resulting in a cumulative improvement of over 700 basis points over these past 6 quarters.

The third quarter marked the second full quarter of regionalization within the U.S. and we're pleased with the early results. Regionalization has allowed us to simplify the network by reducing the number of line-haul lanes, increasing volume within existing line-haul lanes and adding more direct fulfillment center to delivery station connections. We have also been focused on optimizing inventory placement in a new regionalized network which when coupled with the simplification mentioned earlier, is helping contribute to an overall reduction in cost to serve.

Additionally, in the quarter, we saw benefits from lower inflation, primarily within line haul, ocean and rail shipping rates which were partially offset by higher fuel prices. While we are encouraged by the improvements in operating profit, we still see a lot of opportunity in front of us. In international, we were closer to breakeven during the quarter with an operating loss of $95 million. This was an improvement of $2.4 billion year-over-year. This improvement was primarily driven by lowering our cost to serve through higher productivity, decreased inflationary pressures and improvements in leverage across our established and emerging international countries as we continue to focus on customer inputs and improve efficiencies within our operations.

Moving to AWS; revenues were $23.1 billion, an increase of 12% year-over-year. On a quarter-over-quarter basis, we added more than $900 million of revenue in AWS as customers are continuing to shift their focus towards driving innovation and bringing new workloads to the cloud. Similar to what we shared last quarter, while optimization still remain a headwind, we've seen the rate of new cost optimization slowdown in AWS and we are encouraged by the strength of our customer pipeline.

Customers are excited about our approach to generative AI, with several new announcements made during the quarter, including a strategic collaboration with Anthropic, opening Amazon

Bedrock up to general availability, adding Meta's Llama 2 model to Bedrock in the near future and new customization capabilities of CodeWhisperer. AWS remains a clear cloud infrastructure leader with a significant leadership position in the number of customers, the size of our partner ecosystem, our breadth of functionality and the strongest operational performance in the industry.

When we look at the fundamentals of the business, we believe we are in good position to drive future growth as the rates of cost optimization slow down. AWS operating income was $7 billion, an increase of $1.6 billion year-over-year. Our operating margin for the quarter was 30.3%. This is an improvement of approximately 600 basis points quarter-over-quarter, primarily driven by increased leverage on our head count costs.

Shifting to free cash flow. On a trailing 12-month basis, free cash flow adjusted for finance leases was $20.2 billion, an improvement of $41.7 billion year-over-year. The largest driver of the improvement in free cash flow is our increased operating income which we're seeing across all 3 of our segments. Key drivers of this improvement include reductions in our cost to serve, continued advertising growth and improved leverage on our fixed costs. We are also seeing improvements in working capital, notably with our inventory efficiency as we improve our inventory placement.

Now, let's turn to our capital investments. We define our capital investments as a combination of CapEx plus equipment finance leases. These investments were $50 billion for the trailing 12-month period ended September 30, down from $60 billion in the comparable prior year period. For the full year 2023, we expect capital investments to be approximately $50 billion compared to $59 billion in 2022. We expect fulfillment and transportation CapEx to be down year-over-year, partially offset by increased infrastructure CapEx to support growth of our AWS business, including additional investments related to generative AI and large language model efforts.

As we head into the fourth quarter, we are ready to make this a great holiday season for our customers. Looking at our operations network, our inventory is the best position it's ever been heading into the holiday season, enabling us to serve customers with fast delivery speeds from their local regions. We continue to believe putting customers first is the only reliable way to create lasting value for our shareholders.

With that, let's move on to questions.

Question-and-Answer Session

Operator

[Operator Instructions] And the first question comes from the line of Justin Post with Bank of America. Justin, your line is live.

Justin Post

I'll ask about AWS. I guess the first question is, as you look forward in the fourth quarter, you mentioned you've signed some new deals. Are you seeing less cost optimization as you look forward? Or do you think it will be similar to Q3? And then second, you couldn't help but notice the big margin improvement in AWS all the way back to where you were 7 quarters ago. Could you talk about the drivers and sustainability of those margins?

Andrew Jassy

Yes, Justin. Sure. This is Andy. I think if you look at AWS, we grew 12% year-over-year. And I think that you saw a continued stabilization of our year-over-year growth rate. And included $919 million of incremental quarter-over-quarter revenue which it's hard to compare and do the math across different players because not everyone discloses them clearly. But as far as we can tell, that also looks like the most absolute growth of any of the players out there. We're still seeing elevated customer optimization levels than we've seen in the last year or year before this, I should say. So if you look at the optimization, it's more than a year ago but it's meaningfully attenuating from where we've seen in the last few quarters. And if you look at the optimization too, what's interesting is it's not all customers deciding to shut down workloads.

A very significant portion of the optimization are customers taking advantage of enhanced price performance capabilities in AWS and making use of it. So for example, if you look at the growth of customers using EC2 instances that are Graviton based which is our custom chip that we built for generalized CPUs, the number of people and the percentage of instances launched their Graviton base as opposed to Intel or AMD base is very substantially higher than it was before.

And one of the things that customers love about Graviton is that it provides 40% better price performance than the other leading x86 processors. So you see a lot of growth in Graviton. You also see a lot of customers who are moving from the hourly on demand rates for significant portions of their workloads to 1- to 3-year commitments which we call savings plans. So those are just good examples of some of the cost optimization that customers are making in less certain economies where it's really good for customers short and long term and I think it's also good for us. But we're starting to see -- as I said, we're seeing that optimization attenuate. I expect that to continue over time. I remain very optimistic about AWS in the medium to long term. And it's because we have the most functionality by a large amount. We have a much larger partner ecosystem than you'll find elsewhere. We have stronger security and operational performance than you can find elsewhere.

I think we're the most customer-focused. Even if you look during optimization times in this difficult economy, customers have really noticed how AWS leaned in with them for the long term. And I think that matters to customers. And then we have a $92 billion revenue run rate business where 90% of the global IT spend still resides on premises. And if you believe like we do, that equation is going to flip. There's a lot more there for us.

And then you look at the very substantial gigantic new generative AI opportunity which I believe will be tens of billions of dollars of revenue for AWS over the next several years, I think we have

a unique and broad approach that's really resonating with customers. And you can see it with the array of customers I mentioned that are using us and starting to build workloads for generative AI who have already on top of us.

I could see it also. Just the growth rate for us in generative AI is very fast. Again, I have seen a lot of different numbers publicly. It's real hard to measure an apples-to-apples. But in our best estimation, our -- the amount of growth we're seeing and the absolute amount of generative AI business we're seeing compares very favorably with anything else I've seen externally. So I think that you can see that in the deals that we're doing, too.

I spoke about the significant number of deals that we've closed over the last couple of months. A lot of those conversations, it's -- a big piece of it is what I mentioned earlier, the functionality, the ecosystem, the operational performance, the security, how you take care of customers but also whether or not they like your vision and set of products that you're building in this brand-new technology space, generative AI which I think has been very effective so far. So to me, I think there's a lot of growth in front of AWS. I'm very optimistic about it.

Brian Olsavsky

And Justin, on your comment about -- a question about AWS margins. So yes, the margin improved 600 basis points quarter-over-quarter, an increase of income of $1.6 billion quarter-over-quarter for AWS, is driven by -- primarily by our headcount reductions in Q2 and also continued slowness in hiring, rehiring open positions. There's been also a lot of cost control in non-people categories, things like infrastructure costs and also discretionary costs. Natural gas prices and other energy costs have come down a bit in Q3 as well. So as we've said historically, the margins on -- the operating margin for AWS is going to fluctuate quarter-to-quarter and this is a good example of that.

Operator

Our next question is from Doug Anmuth with JPMorgan.

Doug Anmuth

Can you just talk more about how the regional fulfillment network is exceeding your expectations? And then also, how does that support your confidence in moving North America beyond mid-single-digit margin levels? And then just perhaps on generative AI, obviously, a lot of innovation here. You talked about a number of different customers running early workloads. How should we think about the timing, just to drive some tangible monetization there?

Brian Olsavsky

Well, I'll start. The -- on regionalization, I think it was such a significant change for us in the network. It's hard to really appreciate with a big change across so many dimensions it was that first to make such a change has all sorts of risk. And I think the team did a great job planning around it being very thorough. I think that we were able, with our placement algorithms, to get

more regional local in-stock levels than we anticipated. And then I think until you actually put all the connections together, we changed a lot of the connections from a lot of these middle connections going from fulfillment centers to sortation centers and then to delivery stations to connecting the fulfillment centers and delivery stations more directly which, of course, is easier to do when you have more local in stock in a region.

But I think that we were able -- the connections that we made and the optimization that we made there allowed us to get shorter transportation distances than we even anticipated in items to people a lot quicker. And then always one of the issues you got to worry about when you make a change that big is whether or not you end up splitting shipments and having more -- fewer items per box and per shipment.

And that was also something that we really had to work through in the design and in the early stages and we've gotten a lot better on. So we're just across the board, we're seeing shorter transportation distances and much faster delivery to customers. And when you deliver faster delivery to customers, they actually start to consider you for a lot more items than they otherwise would. I think it's part of what -- if you look at our very significant growth rates right now in everyday essentials and consumables, a lot of it is when you -- if you're going to order something which you need in the same day or next day, you're not going to consider it if it's coming in 3 or 4 days. But when you're consistently getting it in same day or the next day, just changes what you're willing to do.

I think the second question was on gen AI and the timing of monetization. What I would tell you is that we have been surprised at the pace of growth in generative AI. Our generative AI business is growing very, very quickly, as I mentioned earlier. And almost by any measure, it's a pretty significant business for us already. And yet I would also say that companies are still in the relatively early stages.

I mean now you have to get perspective. My perspective is that the cloud is still in the early stages. If you think about 90% plus of the global IT spend being on-premises, where I think that equation is going to flip in 10 years. I think cloud is early. So if you -- with that lens on, I still think we're very early in generative AI. And what's interesting, too, around generative AI because we have so many companies who are doing all sorts of prototypes and it's really accelerating very rapidly on the training side with Trainium and Inferentia and then on the application building and running side with Bedrock, is that companies are still trying to sort out for themselves what they're going to run at large-scale production in all of these areas.

Because what happens is you try a model, you test the model, you like the results of the model and then you plug it into your application. And what a lot of companies figure out quickly is that using the really large -- the large models and the large sizes ends up often being more expensive than what they anticipated and what they want to spend on that application. And sometimes too much latency in getting the answers as it shovels through the really large models.

And so customers are experimenting with lots of different types of models and then different model sizes to get the cost and latency characteristics that they need for different use cases. And it's one of the things that I think is so useful about Bedrock is that customers are trying so many variants right now but to have a service that not only lets you leverage lots of third party as well as Amazon large language miles but also lots of different sizes and then makes the transition of moving those workloads easy between them is very advantageous.

Operator

Our next question is from Brian Nowak with Morgan Stanley.

Brian Nowak

Thanks for taking my question, I have two, one on AWS, one on the retail business. On AWS AI Andy, I recognize you have a pretty multipronged AI approach. But just could you talk us through sort of one or two of the early generative AI products where you're seeing the most early demand and interest? And as you talk to customers, are there still hurdles or pain points that you're not quite serving in your product suite you look to solve and innovate on over the next couple of years in AI? And then the second one, you made a lot of steps on the regionalization of warehouses and making it more efficient. Where are you on robotics in the warehouses? And how should we think about the potential impact of that to drive profitability even higher?

Andrew Jassy

On the AI side, I think that if you're looking for some of the products that we're offering that are -- that have a lot of early resonance and traction, I would -- I'd start with Bedrock. These customers are -- they're very excited about Bedrock. And it's making it so much easier to get applications, generative AI applications built. And again, it's machine learning and AI has been something that people have been excited about for 25 years.

In my opinion, about a half dozen years ago, it took a pretty significant leap forward where it was much easier given the economics and scalability of compute and storage and then some of the tools that we built like SageMaker, it was much easier for everyday developers to start to interact with AI. But it just -- it took another meaningful step forward with generative AI but still it's complicated to actually figure out which models you want to work, you want to use and how you actually want to employ them and trying to make sure you have the right results, trying to make sure you get safe results, trying to make sure you end up with a cost structure and a customer experience that you want.

And so it's hard. And customers will like -- there's a certain number of customers who have very deep AI expert practitioners but most companies don't. And so Bedrock just takes so much of the difficulty out of those decisions and those variables that people are very excited about Bedrock. They're using it in a very broad way. They're extremely excited about not just the set of models in there but if you look at a leader like Anthropic and the ability for our customers in Bedrock to have exclusive early access to models and customization tools and fine-tuning which

gives them more control, there's just -- there's a lot of buzz and a lot of usage and a lot of traction around Bedrock.

I would also say our chips, Trainium and Inferentia, as most people know, there's a real shortage right now in the industry in chips. It's really hard to get the amount of GPUs that everybody wants. And so it's just another reason why Trainium and Inferentia are so attractive to people. They have better price performance characteristics than the other options out there but also the fact that you can get access to them. And we've done, I think, a pretty good job providing supply there and ordering meaningfully in advance as well.

And so you're seeing very large LLM providers make big bets on those chips. I think Anthropic deciding to train their future LLM model on Trainium and using Inferentia as well is really a statement. And then you look at the really hot start-up Perplexity.ai, who also just made a decision to do all their training and inference on top of Trainium and Inferentia. So those are 2 examples.

I'd say CodeWhisperer, too, is just, again, it's just a game changer if you can allow your engineers not to have to do the more repetitive work of cutting and pasting and building certain functions that really, if somebody knew your code base better, could do. And so it's a real -- it's a productivity game changer for developers. And then actually launching that customization vehicle so that they actually understand your own proprietary code base, that is something that customers are quite excited about. So those are ultimately early traction AI.

There's so much more to provide, Brian. I mean there's -- I think people -- even though Bedrock is so much easier to use than people trying to build models themselves and build the applications, I think people are still looking to find ways to make it easier to look at a big corpus of data and run an agent on top of it or maybe they don't have to do all that work themselves. I think people are looking for automated ways to understand developer environments and be able to ask any question on the developer environment. So there's a lot more. It's going to be a long time before we run out of services. And yes, I think it's a good thing to look toward the next few months in re:Invent to see the additional things that team launches.

I think on the robotics piece, it's a very significant investment for us. It has been for several years. It's made a huge difference for us. It's made a big difference for us both on the productivity side, on the cost side as well as importantly on the safety side where we can have our teammates working on things that are even safer than what they get to work on today. We have a very substantial investment of additional robotics initiatives, I would say many of which are coming to fruition in 2024 and 2025 that we think will make a further additional impact on the cost and productivity and safety and our fulfillment service.

Operator

Our next question comes from the line of Eric Sheridan with Goldman Sachs.

Eric Sheridan

Maybe one follow-up on AWS and one on the ads business. Andy, would love your perspective. The cloud optimization theme started in the second half of '22 when there were a lot of macro concerns. And then the AI theme really only sort of came to the forefront in the last 8 or 9 months. What's your perspective on how turning the calendar into 2024 and there being a new IT budget cycle could possibly lead us to put the optimization theme in the background and some of the AI theme come more to the forefront when there might be more distinct budgeting around AI as a theme? That would be number one. And then in your ads business, you're approaching $50 billion run rate and it's compounding in the mid-20s. What are you most excited about on the initiative front to continue to build scale both on Amazon properties and possibly off Amazon as a broader digital advertising player?

Andrew Jassy

On the AWS side, my perspective, we'll obviously all have to wait and see to some extent. But my perspective is that in 2024, you're going to -- I think a lot of the relatively low-hanging fruit on optimization has happened in 2023. It's not to say there won't be any more optimization. It's just that there's more low-hanging fruit when you have very large footprints and you've built a lot of applications on a platform for you to go decide to optimize if that's what you want to go do.

And so I think 2020 -- we're already seeing it now with the attenuation of optimization over the last several months. But I think you'll continue to see the attenuation continue and we're already seeing more and more companies that are turning their attention to newer initiatives. And I think what you will see in '24 and '25 as well, I think -- I don't think it's a 1-year deal, I think that's going to be a several-year trajectory, is that you're going to just see a lot of companies not just looking at the new generative AI workloads but also there was a significant number of new customer transformations where companies were going to largely move from being on-premises to being in the cloud.

That got stalled in 2023 because companies were being more conservative with their spend and wary of an uncertain economy. And so I think that what you'll see increasingly is that companies will both go back to those transformations they were planning on making and working with a lot of systems integrator partners as well as ourselves as well as start to see the production in large scale of the generative AI applications that they're all working on and prototyping and starting to deploy into production.

I think on the ad front, well, there's a lot I'm excited about on the ad side as well. I think that it's interesting what's happened in our ads business as if you look around the industry, most advertising-heavy companies have struggled growth-wise as the economy has been difficult. And while we see companies being more cautious on the ad side and the top-of-funnel products, things like display and a little bit of video, we're still seeing a lot of strength in the lower-funnel ad products like sponsored products.

And I think in these types of economies, we have fared pretty well in part because we have a number of owned and operated properties that have very large volumes that advertisers and brands want to get in front of. Even in a harder economy, there's going to be a lot of

e-commerce purchasing. So people want to be in front of our customers in our marketplace. Or take Thursday Night Football which is we're in our second season of Thursday Night Football and off to a great start, the ratings are 25% higher than they were a year ago through 6 weeks. But also, we're doing much better on the advertising side than we did in our first year and that's a property that's really valuable. It's the one game that week and advertisers want to be in front of customers because there's 13 million customers a week watching. So I think part of it is because we have owned and operated properties that have a lot of volume.

And then I think the other piece is that most of our resource in the advertising side is spent on machine learning expert practitioners who are owning algorithms to make sure that the sponsor results people get when they search on something are relevant. And because of that, those ads perform better for advertisers. So when they have to think about budget decisions, they're going to choose the ones that have large volume and perform better. I think both of those are real advantages in our advertising area right now.

In terms of additional things we're excited about. I think that we have barely scraped the surface with respect to figuring out how to intelligently integrate advertising into video, into audio and into grocery. So I think we're early days in that. I think that we also started externalizing some of our products like sponsored products to third-party websites. And you see that with what we've done with Pinterest, Hearst Newspapers and BuzzFeed, so I think there's -- again, we're still pretty early in that area but it's growing well and we're very focused on continuing to in a great customer experience.

Operator

And our final question comes from the line of Mark Mahaney with Evercore ISI.

Mark Mahaney

Okay. On those AWS deals Andy, that you talked about in the September quarter, was there something different about those deals, different industry, different verticals, different geographic markets? Or is it just kind of a resumption of kind of the deal flow that you've had in years past? So that's the first question. And secondly, Brian, international. Is international finally at a point where it can be sustainably profitable going forwards, maybe except for the seasonally challenged March quarter? Have you wrung out enough efficiencies and gotten enough scales? And the oldest markets there are big enough and scalable enough and profitable enough that it offsets the newer countries? Have you finally reached that point?

Brian Olsavsky

First, just to give Andy a break. On international, thanks for your question. Yes, the quarter was just short of breakeven. And as you pointed out, that's a departure from kind of prior trends. I would answer that question this way. There's multiple things going on internationally. In our established countries, U.K., Germany, Japan, France, those countries are profitable and have been profitable. And we continue to work on price selection and convenience. And all that retail base is essentially adding sellers, adding vendors and selection, scaling advertising and

improving the cost structure of our ops network. So many of the ops productivity initiatives, probably with the exception of regionalization, is more of a U.S. side right now. But we're working on all those productivity elements and speed concurrently globally. And you're seeing that internationally and customers are responding.

On the emerging side, over the last 6 years, we've launched 10 new countries. History has shown us that those all take time to grow into profitability. The U.S. took 10 years -- excuse me, 9 years originally. And they're all on their own journey there with growth and scale and profitability and selection and a number of other variables. So that's -- those are all going well as well. We continue -- we're going to continue to invest internationally in things like Prime and expanding Prime benefits and we're going to continue to build out the fulfillment and transportation networks to better serve customers. So I can't say it's permanently we've reached a breakeven threshold for profitability. I think the trend lines are clear, though and we'll continue to work on accelerating that journey in all countries, especially the emerging ones.

Andrew Jassy

On the AWS deals, it's a really broad mix of industries and geographies. So it's not comped up in one. Some are kind of first really big deals from customers. Some of them are very large expansions of existing agreements where they've gone from call it, 20% of their workloads to 50% of their workloads moving to AWS in the cloud. It's also -- I'm not even really including in that number, a number of really big public sector deals that we've done over the last chunk of time that won't hit for a period of time; and so all these deals don't hit in a month. They happen over a period of time as you help those customers safely transition and migrate their workloads to AWS. But it hasn't been any one piece. And I wish I could tell you we knew exactly why they're starting to happen in faster numbers.

I do think, in general, historically, deal volume tends to be lumpy and it doesn't perfectly distribute over a calendar year. I do think though that we've seen this in '23, for sure, that the time to close deals lengthened. And I think it's all reflective of what most companies in the world have been thinking about the last year which is just in the face of uncertain economy, you're going to be more conservative. There are going to be more people involved. You're going to spend more time on how you can save on your existing cost instead of migrating new workloads or thinking about signing new deals.

And so I think what you're just starting to see along with some of the optimization attenuation is that companies are starting to look forward again. And so we've just seen a collection of those the last couple of months that had been -- being discussed for several months where I think, frankly, both sides thought they would close faster but just went slower than they did. So I just think you're starting to see companies look forward more.

Dave Fildes

Thanks for joining us today for the call and for your questions. A replay will be available on our Investor Relations website for at least 3 months. We appreciate your interest in Amazon and look forward to talking with you again next quarter.

Operator

Ladies and gentlemen, that does conclude today's teleconference. You may disconnect your lines at this time. Thank you for your participation.