

Amazon.com, Inc. (NASDAQ:[AMZN](#)) Q3 2024 Earnings Call Transcript October 31, 2024 5:00 PM ET

Company Participants

Dave Fildes - VP of IR
Andy Jassy - CEO
Brian Olsavsky - CFO

Conference Call Participants

Doug Anmuth - JPMorgan
Ross Sandler - Barclays
Brian Nowak - Morgan Stanley
Eric Sheridan - Goldman Sachs
Colin Sebastian - Baird
Justin Post - Bank of America

Operator

Thank you for standing by. Good day, everyone, and welcome to the Amazon.com Third Quarter 2024 Financial Results Teleconference. At this time, all participants are in a listen-only mode. After the presentation, we will conduct a question-and-answer session. Today's call is being recorded. And for opening remarks, I will be turning the call over to the Vice President of Investor Relations, Mr. Dave Fildes. Thank you, sir. Please go ahead.

Dave Fildes

Hello, and welcome to our Q3 2024 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 2023. Our comments and responses to your questions reflect management's views as of today, October 31, 2024 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.

Andy Jassy

Thanks, Dave. Today, we're reporting \$158.9 billion in revenue, up 11% year-over-year excluding the impact from foreign exchange rates. Operating income was \$17.4 billion, up 56% year-over-year, and trailing 12-month free cash flow adjusted for equipment finance leases was \$46.1 billion, up 128% or \$25.9 billion year-over-year. As always, we're focused on making our customers' lives better and easier and thinking long term with respect to how we can keep helping customers and build a successful business to outlast all of us.

In our stores business, we saw sales growth of 9% year-over-year in the North America segment and 12% year-over-year in the international segment. Our team continues to focus on the inputs that matter most to customers, really broad selection, low prices, fast and free delivery, and a range of compelling Prime member benefits, including our recent additions of unlimited grocery delivery from Whole Foods Market, Amazon Fresh, and local third-party grocery partners for \$9.99 a month and fuel savings of \$0.10 a gallon at bp, Amoco, and ampm stations in the US.

At a time when consumers are being careful about how much they spend, we're continuing to lower prices and ship even more quickly, and we can see this resonating with customers as our unit growth continues to be strong and outpace even our revenue growth. In the last few months, we've hosted our largest and most successful Prime Day and Prime Big Deal Days ever and helped customers save over \$5 billion across more than 50 million deals. And for the second year in a row, we are on track to deliver our fastest speeds ever for Prime members globally.

We also continue to focus on lowering our cost to serve and are pursuing several initiatives that we believe will have meaningful long-term impact in this area. First, we continue to believe there are more gains on top of what we've captured thus far in outbound regionalization and getting more items closer to end consumers. As such, we're in the process of significantly changing the way we inbound items into our fulfillment network and subsequently spread them to our regional fulfillment nodes. In the last few months, we've made hundreds of changes to our US inbound network and opened more than 15 inbound buildings. While still relatively early in this re-architecture, we've already improved our ability to spread inventory across our fulfillment

centers by 25% year-over-year, allowing us to have more of the requisite items in fulfillment centers closest to the customer, so we can compile shipments and ship to customers even more quickly. As we scale and optimize this new design, we expect these changes will further improve inventory placement, offer faster delivery time, save transportation costs, and enable us to increase units shipped per box.

Second, we continue to roll out same-day delivery facilities, which is not only the fastest way to get products to customers but also one of our lowest cost ways to deliver. Over 40 million customers this past quarter have had their orders delivered for free with same-day delivery, an increase of more than 25% year-over-year. And third, we continue to innovate in robotics to speed delivery, lower cost to serve, and further improve safety in our fulfillment network. We recently launched our 12th-generation fulfillment center design with the first building launching in Shreveport, Louisiana. This is the first facility that incorporates our newest robotics inventions that simplify stowing, picking, packing, and shipping processes. Thus far, this new design reduces fulfillment processing time by up to 25%, increases the number of items we can offer for same-day or next-day delivery and is expected to drive a 25% improvement in our cost to serve during peak within this next generation facility. Though we believe we have more expansive automation and robotics than other retail peers, it's still early days in how much automation we expect in our fulfillment network.

In advertising, we remain pleased with our progress, generating \$14.3 billion of revenue in the quarter, 18.8% year-over-year growth. Our expansive reach, ability to service relevant offers to our customers, opportunity to engage customers from the top of the funnel to point of purchase, and leading capabilities around measuring outcomes at every touch point provide all types of brands with full funnel advertising at scale.

With sponsored products, we're seeing meaningful growth on a very large base, and we see further opportunity in driving even better performance for advertisers by further improving the relevancy of the ads we show and by providing additional optimization controls. At the same time, some of our newer offerings are in their very early days. We're just entering our first broadcast season for Prime Video advertising, following a very strong showing at upfront. And we're continuing to support brands of all sizes with our Generative AI-powered creative tools across display, video and audio, including our video generator that uses a single product image to curate custom AI-generated videos. While we're generating a lot of advertising revenue today, there remains considerable upside.

AWS grew 19.1% year-over-year and now stands at a \$110 billion annualized run rate. We've seen significant reacceleration of AWS growth for the last four quarters. With the broadest functionality, the strongest security and operational performance and the deepest partner community, AWS continues to be a customer's partner of choice. There are signs of this in every part of AWS's business. We see more enterprises growing their footprint in the cloud, evidenced in part by recent customer deals with the ANZ Banking Group, Booking.com, Capital One, Fast Retailing, Itaú Unibanco, National Australia Bank, Sony, T-Mobile, and Toyota.

You can look at our partnership with NVIDIA called Project Ceiba, where NVIDIA has chosen AWS's infrastructure for its R&D supercomputer due in part to AWS's leading operational performance and security. And you can see how AWS continues to innovate in its infrastructure capabilities. With deliveries like Aurora Limitless Database, which extends AWS's very successful relational database to support millions of database writes per second and manage petabytes of data whilst maintaining the simplicity of operating a single database or with our custom Graviton4 CPU instances, which provide up to nearly 40% better price performance versus other leading x86 processors.

Companies are focused on new efforts again, spending energy on modernizing their infrastructure from on-premises to the cloud. This modernization enables companies to save money, innovate more quickly, and get more productivity from their scarce engineering resources. However, it also allows them to organize their data in the right architecture and environment to do Generative AI at scale. It's much harder to be successful and competitive in Generative AI if your data is not in the cloud.

The AWS team continues to make rapid progress in delivering AI capabilities for customers in building a substantial AI business. In the last 18 months, AWS has released nearly twice as many machine learning and GenAI features as the other leading cloud providers combined. AWS's AI business is a multibillion-dollar revenue run rate business that continues to grow at a triple-digit year-over-year percentage and is growing more than 3 times faster at this stage of its evolution as AWS itself grew, and we felt like AWS grew pretty quickly.

We talk about our AI offering as three macro layers of the stack, with each layer being a giant opportunity and each is progressing rapidly. At the bottom layer, which is for model builders, we were the first major cloud provider to offer NVIDIA's H200 GPUs through our EC2 P5e instances. And thanks to our networking innovations like Elastic Fabric adapter and Nitro, we continue to offer advantaged networking performance. And while we have a deep partnership with NVIDIA, we've also heard from customers that they want better price performance on their AI workloads. As customers approach higher scale in their implementations, they realize quickly that AI can get costly. It's why we've invested in our own custom silicon in Trainium for training and Inferentia for inference. The second version of Trainium, Trainium2 is starting to ramp up in the next few weeks and will be very compelling for customers on price performance. We're seeing significant interest in these chips, and we've gone back to our manufacturing partners multiple times to produce much more than we'd originally planned.

We also continue to see increasingly more model builders standardize on Amazon SageMaker, our service that makes it much easier to manage your AI data, build models, experiment, and deploy to production. This team continues to add features at a rapid clip punctuated by SageMaker's unique hyperpod capability, which automatically splits training workloads across more than 1,000 AI accelerators, prevents interruptions by periodically saving checkpoints, and automatically repairing faulty instances from their last saved checkpoint and saving training time by up to 40%. At the middle layer where teams want to leverage an existing foundation model customized with their data and then have features to deploy high-quality Generative AI

applications, Amazon Bedrock has the broadest selection of leading foundation models and most compelling modules for key capabilities like model valuation, guardrails, rag and agents.

Recently, we've added Anthropic's Claude 3.5 Sonnet model, Meta's Llama 3.2 models, Mistral's Large 2 models and multiple stability AI models. We also continue to see teams use multiple model types from different model providers and multiple model sizes in the same application. There's mucking orchestration required to make this happen. And part of what makes Bedrock so appealing to customers and why it has so much traction is that Bedrock makes this much easier. Customers have many other requests, access to even more models, making prompt management easier, further optimizing inference costs, and our Bedrock team is hard at work making this happen.

At the application or top layer, we're continuing to see strong adoption of Amazon Q, the most capable Generative AI-powered assistant for software development and to leverage your own data. Q has the highest reported code acceptance rates in the industry for multiline code suggestions. The team has added all sorts of capabilities in the last few months, but the very practical use case recently shared where Q Transform saved Amazon's teams \$260 million and 4,500 developer years in migrating over 30,000 applications to new versions of the Java JDK as excited developers and prompted them to ask how else we could help them with tedious and painful transformations. Spend a few minutes reading developer forums about what they wish they could move away from, and you'll get an idea of what they want. Expect more practical AI game changers from Q.

We're also using Generative AI pervasively across Amazon's other businesses with hundreds of apps in development or launched. For consumers, we've expanded Rufus, our Generative AI-powered expert shopping assistant to the UK, India, Germany, France, Italy, Spain, and Canada. And in the US, we've added more personalization, the ability to better narrow customer intent and real-time pricing and deal information. We've recently debuted AI Shopping Guides for consumers, which simplifies product research by using Generative AI to pair key factors to consider in a product category with Amazon's wide selection, making it easier for customers to find the right product for their needs.

For sellers, we've recently launched Project Amelia, an AI system that offers tailored business insights to boost productivity and drive seller growth. We continue to rearchitect the brain of Alexa with a new set of foundation models that we'll share with customers in the near future, and we're increasingly adding more AI into all of our devices. Take the new Kindle Scribe we just announced. The note-taking experience is much more powerful with the new built-in AI-powered notebook, which enables you to quickly summarize pages of notes into concise bullets in a script font that can easily be shared.

Speaking of Kindle products, we just launched a completely new Kindle lineup, the first time we've done a portfolio refresh of this size. Apart from the Scribe, it includes the first-ever color Kindle, the fastest Kindle Paperwhite ever and a new pocket-sized Kindle. The early sales for these devices have significantly outperformed our expectations, and Kindle is having an

excellent year with customers reading more than ever. We now have over 20 billion average monthly pages read on Kindle devices worldwide.

There are so many things we're energized by right now, but we'll quickly mention one more, the progress we're making in improving customers' pharmacy experience. Brick-and-mortar pharmacies account for just over 90% of prescriptions dispensed in the US that require customers to make trips to forlorn physical venues with much of the selection behind locked shelves, waiting lines for meds and only finding out about pricing at the point of purchase. The largest mail order pharmacies offer delivery in 5 to 10 business days. We think customers deserve better.

Today, we can deliver to 95% of first-time Amazon Pharmacy customers in the US within two business days and to 20% of US Prime members within 24 hours. Next year, we plan to launch operations in 20 new cities, so nearly half the US will have the ability to have their medications delivered to their door within hours. We believe making it easier for customers to get their medications will improve medication adherence, which we know can directly improve health outcomes. Across Amazon, we have a lot coming in the next few months. We eagerly look forward to sharing these with customers, and a big thank you to my Amazon teammates around the world for all their hard work.

And with that, I'll turn it over to Brian for a financial update.

Brian Olsavsky

Thanks, Andy. Let's start with our top line financial results. Worldwide revenue was \$158.9 billion, an 11% increase year-over-year, excluding a 20 basis point unfavorable impact of foreign exchange. As a reminder, our Q3 guidance had anticipated a larger unfavorable impact of approximately 90 basis points. Worldwide operating income increased 56% year-over-year to \$17.4 billion, our highest quarterly operating income ever and was \$2.4 billion above the high end of our guidance range. We remain focused on streamlining and managing costs in a way that allows us to continue inventing for customers in a cost-effective way.

In the third quarter, North America segment revenue was \$95.5 billion, an increase of 9% year-over-year. International segment revenue was \$35.9 billion, an increase of 12% year-over-year. Worldwide paid units accelerated to 12% growth year-over-year as our customers continue to come to Amazon for low prices, broad selection and convenient fast delivery. Prime remains a core contributor to this growth. Year-over-year paid membership growth accelerated in Q3 from both the US and globally, helped by our tenth annual Prime Day event in mid-July.

Customers are enjoying even faster delivery speeds, which also helps drive strong growth in items like everyday essentials. This includes items like health, beauty, and personal care as well as nonperishable grocery. Though all these items often have a lower average selling price, the strength in everyday essential's revenue is a positive indicator that customers are turning to us for more of their daily needs. We see that when customers purchase these types of items from

us, they build bigger baskets, shop more frequently and spend more on Amazon. We remain focused on keeping prices sharp and offering broad selection that gives customers options when making purchase decisions.

Shifting to profitability. Our North America and international segments each delivered their seventh consecutive quarter of year-over-year operating margin improvement. North America segment operating income was \$5.7 billion, an increase of \$1.4 billion year-over-year. North America operating margin was 5.9%, up 100 basis points year-over-year. In the third quarter, we continue to make progress in improving our fulfillment network cost structure, driven in part by improved inventory placement. This helped us drive better productivity in our transportation network and shipping efficiency from higher units per box.

In the international segment, operating income was \$1.3 billion, an improvement of \$1.4 billion year-over-year. International operating margin was 3.6%, up 390 basis points year-over-year. So far this year, we've generated operating profit in each of the three quarters for the international segment and totals \$2.5 billion year-to-date. We're seeing strength in our established countries like the UK and Germany as we continue to drive efficiencies to improve on-road productivity in our transportation network and better execution in our fulfillment centers.

Our emerging countries are growing revenue at a healthy rate also, and they are leveraging their cost structures on investing strategically in Prime benefits. We have confidence that our focus on the inputs, coupled with the strength of our global teams, will continue to drive improvement over time.

Advertising remains an important contributor to profitability in the North America and international segments. This quarter, we saw strong growth on an increasing large base of advertising revenue. We see many opportunities to further expand our ads offering in areas that are driving growth today like Sponsored Products as well as more recent growth areas like Prime Video ads.

Moving next to our AWS segment. Revenue was \$27.5 billion, an increase of 19.1% year-over-year, excluding the impact of foreign exchange. AWS now has an annualized revenue run rate of \$110 billion. The business continues to grow, mostly opportunity to expand our core cloud offering and our AI services. Customers increasingly recognize that to get the true benefit of Generative AI, they also need to move to the cloud.

AWS operating income was \$10.4 billion, an increase of \$3.5 billion year-over-year as a result of our continued focus on cost control, including a measured pace of hiring, a focus on driving efficiencies in our infrastructure, and reducing costs across the business. Additionally, we increased the estimated useful life of our servers starting in 2024, which contributed approximately 200 basis points to the AWS margin increase year-over-year in Q3. As we said in the past, we expect the AWS operating margins to fluctuate, driven in part by the level of investments we're making at any point in time.

Now turning to our capital investments. As a reminder, we define these as a combination of cash CapEx plus equipment finance leases. Year-to-date capital investments were \$51.9 billion. We expect to spend approximately \$75 billion in CapEx in 2024. The majority of the spend is to support the growing need for technology infrastructure. This primarily relates to AWS as we invest to support demand for our AI services while also including technology infrastructure to support our North America and international segments. Additionally, we're continuing to invest in our fulfillment and transportation network to support the growth of the business, improve delivery speeds and lower our cost to serve. This includes investments in same-day delivery facilities, in our inbound network and as well in robotics and automation. We're encouraged by the start of the holiday season, which kicked off in October with a strong Prime Big Deal Days.

We are ready to serve customers throughout the season, and I want to thank our teams across Amazon for delivering two very large Prime member events in the past four months and for getting us ready to delight customers during this holiday season.

With that, let's move on to your questions.

Question-and-Answer Session

Operator

[Operator Instructions] And the first question comes from the line of Doug Anmuth with JPMorgan. Please proceed with your question.

Doug Anmuth

Thanks so much for taking the questions. Brian, I was hoping you could talk a little bit more about the drivers of the 38% AWS margins. I know you mentioned the 200 basis points related to the useful lives push-out, but just how to think about sustainability kind of in the 30s there going forward? And then perhaps related the step-up in CapEx that you saw in 3Q, you gave the full-year number, which was helpful. Any kind of early read or thoughts on how we should think about 2025? Thanks so much.

Brian Olsavsky

Okay. Thanks, Doug. Let me start with AWS. Yeah, the primary drivers of the year-over-year margin increase are threefold. So first is accelerating top-line demand. That helps with all of our efficiencies, our cost-control efforts and of course, what you just mentioned, I believe, the change in the useful life of our servers this year. Let me remind you on that one. We made the change in 2024 to extend the useful life of our servers. This added about 200 basis-points of margin year-over-year. On cost-control, it's been a number of areas. It's -- first is hiring and staffing. We're being very measured in our hiring and you can tell as a company where our office staff is down slightly year-over-year and it's flat to the end of last year. So we're really working hard to maintain our efficiencies not only in the sales force and other areas and production teams, but also in our infrastructure areas. Infrastructure is a big part of our cost structure in AWS and we're probably the best at matching up supply and demand and driving

cost efficiencies and driving those operations very, very strongly. And I think you saw that this quarter, especially on the higher-volume. And over-time, this margin will fluctuate in the segment. It's a number of factors to keep in mind, including our level of investment in the amount of innovation and product development we're doing for new products and services, changes in our staffing for sales and support organizations and of course, capital to support customer growth and to build cost performance chips for our customers' applications. So as I mentioned, we do a really good job of matching supply and demand and we do that in this business, it can result in quarters that we just saw.

Andy Jassy

Yeah, I'll take the capex part of that. As Brian said in his opening comments, we expect to spend about \$75 billion in 2024. I suspect we'll spend more than that in 2025. And the majority of it is for AWS and specifically, the increased bumps here are really driven by Generative AI. As I was mentioning in my own opening comments, our AI business is a multi-billion dollar business that's growing triple-digit percentages year-over-year and is growing three times faster at its stage of evolution than AWS did itself. We thought AWS grew pretty fast. And so the thing to remember about the AWS business is the cash life cycle is such that the faster we grow demand, the faster we have to invest capital in data centers and networking gear and hardware. And of course, in the hardware of AI, the accelerators or the chips are more expensive than the CPU hardware. And so we invest in all of that upfront in advance of when we can monetize it with customers using the resources. But of course, a lot of these assets are many year useful life assets. Data centers, for instance, are useful assets for 20 to 30 years. And so I think we've proven over time that we can drive enough operating income and free-cash flow to make this a very successful return on invested capital business and we expect the same thing will happen here with Generative AI. It is a really unusually large maybe once in a lifetime type of opportunity. And I think our customers, the business and our shareholders will feel good about this long-term that we're aggressively pursuing it.

Operator

And our next question comes from the line of Ross Sandler with Barclays. Please proceed with your question.

Ross Sandler

Andy, if I could push on that last answer, you talked about how AI is about the same size but growing way faster than early AWS. And if we look back at the market in like the early 2010s is hypercompetitive on pricing, your margins were below 15%, I believe, back then. And it seems like a lot of that is what's going on right now in AI with these new AI data centers where you've got competitive pricing and suboptimal utilization just in the whole industry. So as that revenue line grows from where it is today to tens of billions of dollars in coming years, how does that margin come on from the AI data centers versus your existing kind of 30-plus-percent margin at AWS? Any thoughts on how quickly you can close the gap and how that looks on the new AI

workloads versus maybe this mid-30s that the core business is running at? Thank you very much.

Andy Jassy

Yeah. So some quick context, Ross. I think one of the least understood parts about AWS, over time, is that it is a massive logistics challenge. If you think about, we have 35 or so regions around the world, which is an area of the world where we have multiple data centers, and then probably about 130 availability zone through data centers, and then we have thousands of SKUs we have to land in all those facilities. And if you land too little of them, you end up with shortages, which end up in outages for customers. So most don't end up with too little, they end up with too much. And if you end up with too much, the economics are woefully inefficient. And I think you can see from our economics that we've done a pretty good job over time at managing those types of logistics and capacity. And it's meant that we've had to develop very sophisticated models in anticipating how much capacity we need, where, in which SKUs and units. And so I think that the AI space is, for sure, earlier stage, more fluid and dynamic than our non-AI part of AWS. But it's also true that people aren't showing up for 30,000 chips in a day. They're planning in advance. So we have very significant demand signals giving us an idea about how much we need. And I think that one of the differences if you were able to get inside of the economics of the different types of providers here is how well they manage that utilization and that capacity. It has a very direct impact on what kind of margins you have over time and what kind of capital efficiency you also have over time. And so I think you're right, Ross, that there are some similarities in the early days here of AI, where the offerings are new and people are very excited about it. It's moving very quickly and the margins are lower than what they -- I think they will be over time. The same was true with AWS. If you looked at our margins around the time you were citing, in 2010, they were pretty different than they are now. I think as the market matures over time, there are going to be very healthy margins here in the Generative AI space.

Operator

And the next question comes from the line of Brian Nowak with Morgan Stanley. Please proceed with your question.

Brian Nowak

Thanks for taking my questions. Andy, I have two. The first one, I feel like we talk a lot about your targets and sort of north stars for domestic profitability. Can you talk to us a little bit about how you think about the drivers of international retail profitability from here and sort of how you -- just conceptually where the margins could go and what drives them there? And then the second one, I wanted to sort of ask you more about the robotics that you mentioned. I know you've been investing in robotics now for quite a few years. Where are you now in that journey? And how do we think about the next largest areas of investment in robotics in the warehouse network? Thanks.

Brian Olsavsky

This is Brian. Let me take the first part of that question on international segment profitability. So if you've looked at the trends, we see quarter-to-quarter fluctuations certainly in operating profit in the international segment. But we're starting to see clear trend lines that are moving positive and above zero. And for each of the last seven quarters, as I mentioned, we've had year-over-year improvement in op margin. And in the last quarter, operating income was up \$1.4 billion versus the prior year. So really, it is a lot of the same factors you're seeing in North America, lower cost to serve, greater contribution from advertising, improved selection, faster delivery speeds, which help drive consumer demand. And if you step back further, international is really a mix of our established countries where we've been operating for a long time, the UK, Europe, and Germany -- excuse me, Europe and Japan, and a number of emerging countries, including 10 new countries that we launched in the last seven years. So there's a lot under the hood there and there's different stories in each country. Each of them is on a different point in the path from launching to customer adoption, just scaling towards profitability to then making consistent operating profit. So I think our expectations in each of those countries mirror those in North America. And we intend, over the fullness of time, that we're going to aim for North America margins, which are not static themselves. So we're happy with the performance. Glad to see a number above zero in many of the last few quarters, but really it's a story on a country-by-country basis that we're working hard. On the robotics piece, what I would say is even though we believe we have more expansive and advanced automation, robotics capabilities in our fulfillment network than other peers, it's so early with respect to what we're going to do automation, robotics-wise in our fulfillment network. We're just at the stage right now where we're starting to roll out. We had about a five or six very significant new robotics capabilities in the areas of stowing, picking, packing, and shipping that we are finally put into one facility to get the entire workflow. It's a facility in Shreveport, Louisiana that was just launched a few weeks ago. And as I mentioned in my opening comments, we're seeing very encouraging results there. And of course, the reason why we're trying to have more robotics and automation in our fulfillment network is it allows us to fast -- to ship more quickly, to ship more cost effectively, and to make conditions even safer for our fulfillment teammates than what they already have today. And I think something that the team has done in a very disciplined way, and I talked about this a little bit in my annual letter, my shareholder letter is just they have been very thoughtful about defining what are the primitive foundational building blocks that you need, that you can then use in lots of different combinations to build additional automation/robotics capabilities down the road, so we can move even more quickly with the next generation of robotics capabilities. And that team has invested in those and they're already working on the next generation. And the last piece I would add to that is we really do believe that AI is going to be a big piece of what we do in our robotics network. We had a number of efforts going on there. We just hired a number of people from an incredibly strong robotics AI organization. And I think that will be a very central part of what we do moving forward, too.

Operator

And the next question comes from the line of Eric Sheridan with Goldman Sachs. Please proceed with your question.

Eric Sheridan

Thanks for taking the question. Maybe a two-parter, if I can. Looking at the results in the commerce business, there's a clear sort of pattern emerging with respect to ASPs versus unit growth. Can you talk a little bit about what you're seeing in terms of consumer behavior and the propensity to be receptive to either lower ASP items or discounts among consumer behavior? And then the second part would be, can you talk a little bit about some of the strategic initiatives you're building for the long term around capitalizing on lower ASP items and the potential shift to more consumables and the base of consumption habits on the platform? Thanks so much.

Andy Jassy

Yeah, hi, Eric, thank you for your question. Yeah, you're right. Q3 unit volume was strong, 12% worldwide. It's pretty similar in North America versus international. And the consumer factors, we've seen a continuation of many of the things we've discussed over the first -- or last few quarters. Customers looking for deals and are price conscious. This matches up well with our Prime events where we were -- which were well received and saved billions for our -- billions of dollars for our Prime members. It also led to Prime member -- paid Prime member growth accelerating in the quarter, which is hugely impactful for us. A lot of our best offers and deliveries are tuned to Prime membership, and it's the best deal in retail. So with all the benefits you get for the subscription price, so those are all positive signs. When you look at the split between revenue growth and unit growth, you do see some impact of the lower ASP products that we're selling as well as some of the trade down that consumers are doing. But we take that as a real positive, seeing the growth in everyday essential categories, which are really predicated on speed. So it's -- you have to have fast delivery to be able to sell the those products to customers. And when you do, it results in a stickier consumer relationship, higher orders, building larger baskets, which help our ship economics, and repeat orders are stronger. So those are all positive signs, and we'll take any short-term degradation in ASP because what we're focused on primarily is free cash flow here. And we see the ability to unlock all new elements of the consumer spend.

Brian Olsavsky

I'll just add, Eric, on the lower ASP strategy. There's an expression that we've used a lot of times over the years that it's easy to lower prices but it's much harder to be able to afford to lower prices. And the same thing is probably true about lower ASP items. It's pretty easy to choose to supply them but it's much harder to be able to afford to economically supply them. And so one of the reasons that we have been so maniacal about cost to serve over the last few years is that as we're able to take our cost to serve down, it just opens up the aperture for more items, particularly lower ASP items that we're able to supply in an economic way. And when you layer on top of that the broader selection, you layer on top of that faster speed and what we're able to do in getting items and fulfillment nodes local and close to our end customers, what we're able to do in expanding our same-day network, which allows us to ship same-day but also is one of our lowest-cost ways is it just lets us serve so many more of customers' shopping missions. And

that's what we're seeing in our business right now. But again, I think there's a lot of opportunity to continue to expand that.

Operator

And the next question comes from the line of Colin Sebastian with Baird. Please proceed with your question.

Colin Sebastian

Great. Thanks guys. Good afternoon. A question on online stores. The third-party unit mix declined a little bit in Q3, which is unusual. So just wondering if there are any specific drivers of that shift. And then secondly, you made a lot of progress with the AI agents on AWS and for sellers and Rufus for buyers. I wonder with all the underlying data that you have and leveraging those agent capabilities if you could provide any perspective on what a next-generation Alexa might look like and your opportunities there to perhaps drive incremental revenues. Thank you.

Andy Jassy

Sure, Colin. Thanks for your question. On the first one, third-party sellers, yes, the percent of paid units was 60% in Q3. And if you look over the last two years, it's -- essentially, it was at 59%, stepping up to 60%, stepping up to 61% pretty much every second quarter. So what we're seeing is still in a tight band between 59% and 61%. But what we're seeing is the impact of the everyday essentials. There's -- 3P demand is still strong and unit volumes are strong. It's just the everyday essentials tend to skew more to 1P than 3P. And on your second question, Colin, around Alexa, I think we have a really broad number of Alexa devices all over people's homes and offices and automobiles and hospitality suites. We've about 0.5 billion devices out there with a couple of hundred million active endpoints. And when we first were pursuing Alexa, we had this vision of it being the world's best personal assistant and people thought that was kind of a crazy idea. And I think if you look at what's happened in Generative AI over the last couple of years, I think you're kind of missing the boat if you don't believe that's going to happen. It absolutely is going to happen. So we have a really broad footprint where we believe if we rearchitect the brains of Alexa with next-generation foundational models, which we're in the process of doing, we have an opportunity to be the leader in that space. And I think if you look at a lot of the applications today that use Generative AI, there's a large number of them that are having success in cost avoidance and productivity. And then you're increasingly seeing more applications have success in really impacting the customer experience and being really good at taking large corpuses of data and being able to summarize and aggregate and answer questions, but not that many yet that are really good on top of that in taking actions for customers. And I think that the next generation of these assistants and the Generative AI applications will be better at not just answering questions and summarizing the indexing and aggregating data, but also taking actions. And you can imagine us being pretty good at that with Alexa.

Operator

And our final question comes from the line of Justin Post with Bank of America. Please proceed with your question.

Justin Post

Great. Thank you. Maybe a couple of questions. Just on the cloud, are you at all capacity constrained, and will the new Trainium or NVIDIA chips maybe even drive sales growth faster? And then secondly and I know there's been some competition from some of the more traditional retailers growing quickly online. Maybe talk about maybe the advantages that Amazon has with a fulfillment center distribution versus store distribution? Any thoughts on that? Thank you.

Andy Jassy

Well, on cloud, Justin, what I would tell you is that we -- I believe we have more demand than we could fulfill if we had even more capacity today. I think pretty much everyone today has less capacity than they have demand for, and it's really primarily chips that are the area where companies could use more supply. And so we have -- we're growing at a very rapid rate and have grown a pretty big business here in the AI space and it's early days, but I actually believe that the rate of growth there has a chance to improve over time as we have bigger and bigger capacity. And I think that one of the reasons I mentioned in my opening comments that we have a very deep partnership with NVIDIA. We tend to be their lead partner on most of their new chips. We were the first to offer H200s in EC2 instances. And I expect us to have a partnership for a very long time that matters. It's also true that for customers that start to scale out their implementations on the inference side, particularly, they realize pretty quickly that it can get costly. And it's really why we have pursued building Trainium and Inferentia, which is our custom silicon. And the second version of Trainium, Trainium2 will start to ramp up in the next few weeks. And I think it's going to be very compelling for customers on a price performance basis. And we have a lot of customer interest. We have gone back to our manufacturing partners a couple of times now to produce a lot more Trainium than we anticipated. Some of that, for sure, is due to the fact that we just -- we have very large demand, and we want more capacity and supply to be able to provide them. But a lot of it is that customers are excited about the price performance that they believe they're going to get in Trainium. So I think it's a very significant opportunity for us in the AWS space, and it's going to be a very significant opportunity for customers. I think on the retail side, I'd start by just saying that we've always believed that competition is very positive and very healthy. It's good for consumers, good for businesses. It's actually great for innovation. And if you look at the retail space, it's a very, very large market segment. I mean, we have a pretty big retail business, and yet we're only about 1% of the market segment share of the worldwide global retail market segment. And still about 80% to 85% of that market segment share lives in physical stores. And so if you believe that equation is going to flip in the next 10 to 20 years, which we do, there's just a lot of opportunity not just for us but for several players. There won't be only one successful player. I do think that we have some elements of our customer experience that are really unusual and unlike others. I think we have meaningfully broader selection than almost all the players that you probably have seen and heard of. We have low prices with very significant deals that we go work with our third-party selling partners around key holiday shopping occasions. And then we have very significant

advantages and speed of delivery to customers. And we just continue to see in every bit of testing and analysis that we do, that the faster that we're able to promise customers that we can get them their items, the more frequently they buy and the more they actually use Amazon for their shopping needs. And so those are things that are pretty different. I also think we have a way of prioritizing customers that I think is unusual. Our orientation, our DNA and our core at Amazon starts with the customer, and everything moves backwards from that. Any meeting you go to inside of Amazon, we're always asking ourselves what do customers want? What do customers say? What do they not like about the experience? What could be better? And that customer orientation is very important, not just in how you take care of customers, but the world changes quickly all the time. And the technology is changing really quickly. And so if you have the combination of strong technical aptitude, propensity and a passion for inventing and then also a customer orientation where it drives everything you do, I think you have an opportunity to continue to build great trust in a business over a long period of time.

Dave Fildes

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

Operator

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