

Microsoft Ignite

AI transformation for your organization with the Microsoft Cloud

Scott Guthrie; Charles Lamanna; Arun Ulagaratchagan; Ali Ghodsi; Eric Boyd;

Erin Chapple; Rani Borkar; Sarah Bird; Seth Juarez

Thursday, November 16, 2023

SCOTT GUTHRIE: Good morning, everyone. It's so great to be back here at Microsoft Ignite. Our mission at Microsoft is to empower every person and every organization on the planet to achieve more. And with Microsoft Cloud, we're doing this in two ways.

First, we're going to help you radically accelerate employee productivity and capability by building AI and Microsoft Copilot into every Microsoft Cloud solution. Second, we're delivering the tools and services that you need to build transformational AI apps and solutions of your own using Azure.

Yesterday, Satya and Rajesh covered Microsoft Copilot and the Copilot stack. Today, we're going to talk about Azure and the application data and AI platform within it that you can use to build your own Copilots to further drive your AI transformation.

Let's start by talking about the development tools you can use to do it. Microsoft was founded as a developer tools company, and developers remains the very center of our mission. The Visual Studio family of products, which include both Visual Studio and Visual Studio Code, are now the most widely used development tools in the world.

GitHub is used by over 100 million developers and is literally the home of open source. And the Microsoft Power Platform is most widely adopted low-code platform and it helps organizations further accelerate the solutions that they can build.

Their GitHub now provides developers and enterprises an integrated, end-to-end developer platform that can be used to collaborate, automate and secure DevOps solutions. With GitHub, you can build amazing AI solutions and transform your businesses.

One of those exciting capabilities that we've launched with GitHub is a service we call GitHub Copilot. GitHub Copilot was really the first solution that we built using the transformational, large AI models developed by OpenAI. GitHub Copilot provides an AI pair programmer that works with all the popular programming languages, and dramatically accelerates developer productivity.

This month, we're releasing several great new capabilities and features to GitHub Copilot and making GitHub Copilot Chat generally available.

We're also introducing a new subscription tier called GitHub Copilot Enterprise that enhances the entirety of the GitHub Copilot offerings with additional enterprise

capabilities, including the ability to personalize all recommendations using the context of an organization's existing code bases.

Let's take a look at GitHub Copilot and some of the amazing productivity features it now enables.

VIDEO SEGMENT: GitHub Copilot is a mature and trusted AI pair programmer, now used by more than a million developers. It's fantastic at turning natural language comments to code and predicting the code you need next. For example, if I want to add a new route to my node application, Copilot can help me do that. I'll write a comment explaining what I want, and GitHub Copilot writes the code for me following the patterns I've used elsewhere in my file. If it doesn't give me exactly what I want the first time, it will work with me, ultimately making me more productive. And with GitHub Copilot Chat, it truly feels like I'm working with a pair programmer.

For example, I can ask general programming questions like what password standards should I check for in my code? In this case, it even provided some sample code I could use. I have some password validation code already, but it's a regular expression and those are not easy to interpret at a glance. So I can use the "slash explain" command in GitHub Copilot Chat to help explain code I'm unfamiliar with, even if it's a regular expression. I can even ask specific questions about the code I'm looking at. In this case, I have a suspicion we're generating invalid passwords. Sure enough, we are, and Copilot even gives me a suggested fix.

But I want to show you something new. GitHub Copilot Chat can also be invoked inline, so it can work with the code in place, making refactoring simple. To get to it, highlight or put your cursor over the relevant code and hit Ctrl or Command-I. For example, I can ask for a change in natural language and it will show me the changes it's going to make just like a git diff.

Let's accept that change. Copilot Chat inline is also useful for documentation. Using the DOC command, it will add comments specific to the language I'm working in. And even better, I can even ask for tests for the code I'm looking at. Even though I'm using Copilot Chat inline, it's smart enough to know that I need them in another file. So when I accept the changes, it creates one for me.

But like me, GitHub Copilot doesn't just live in my editor, it's also with me once I've finished my work and submitted a pull request. GitHub Copilot for Pull Requests can provide a description of my changes. And best of all, it doesn't just list the changes, it gives useful, meaningful descriptions of what the change actually does. I even get deep links to the lines of code that have changed, if the code reviewers want to drill in deeper.

We're also bringing the power of AI to GitHub Advanced Security. Now, GitHub can not only detect security issues in your code, but it can suggest a fix as well. You can apply the fix to your pull request with one click, or even start a code space with the fix applied if you want to examine it more closely.

GitHub is the AI-powered developer platform that helps your teams write more secure, better documented, and better tested code faster than ever before.

(Applause)

SCOTT GUTHRIE: GitHub Copilot dramatically accelerates developer productivity. 46% of all lines of code written by developers using GitHub Copilot are now fully AI-generated, and 96% of developers report being able to complete repetitive tasks much faster. 75% of developers say that GitHub Copilot helps them focus on more satisfying work and enjoy their jobs more.

I love this last statistic that GitHub Copilot doesn't just bring productivity, it also brings meaning and satisfaction the most. That's the real power and potential of what AI provides.

With Visual Studio and GitHub, we have the most widely-used professional developer tools in the world. With the Microsoft Power Platform, we also have the most widely-used low-code development tools in the world as well.

What I'd like to do now is invite Charles Lamanna, who leads our Dynamics 365 and Power Platform Engineering Team, onto the stage to tell you more about them. Please welcome Charles.

CHARLES LAMANNA: Thank you, Scott.

(Applause)

One of the things which is really incredible is that it's not just about Copilot making you faster with code, but also how Copilot can start to show up using things like low code in the Power Platform. As Scott mentioned, it's the world's most adopted low-code offering for building applications, analyzing data, automating tasks and more.

One of the things we were incredibly excited to announce yesterday was Microsoft Copilot Studio. Microsoft Copilot Studio makes it possible to customize and extend the Microsoft Copilot wherever it shows up, but also to go create your own copilots, drawing from all of the amazing low-code capabilities that we have.

What Copilot Studio can do is it can take advantage of over 1,100 prebuilt plugins and connectors to connect to any of your data sources and orchestrate any of your existing workflows. It also supports all the OpenAI standard concepts like plugins and GPTs and makes it easy for you to make your web app or mobile app or more to Copilot-enabled.

For the demo today, I want to take us through and actually build something inside of Copilot Studio. If you look right here, we can see this is the Copilot that you saw yesterday inside of Microsoft Teams. Like we're all used to, we want to be able to just

converse and talk to Copilot to answer our questions and understand what's going on in our business.

In my job, I work with supply chain and supplier, so I'm going to go see just what's the supplier lead time for one of our common suppliers, Alpine Ski House. I get back a response, and the reality is Copilot doesn't know how to get that information because it's not inside the Microsoft Graph, and it's not available on the web. But the good news is there's a great extensibility story here.

If I click over, I can see that I can go register plugins, but I can also use Copilot Studio to create a plugin for my system if the plugin doesn't already exist. If I go launch this experience, I land right into Copilot Studio, where I can start to build what we call these conversational plugins. What conversational plugins do is they allow me to register a trigger phrase which will execute a series of steps, or even multi-turn conversations based on what the user has entered.

I'm just going to go in and add this conversational plugin, and this makes it really easy to go extend Copilot however I need to do it for my business. Say if I click through here, I land right into a visual designer that everybody knows and loves from the Power Platform. Inside this designer, the first thing I do is I create what's called a trigger. With a trigger, I provide details and prompts about when I want to execute this plugin. This is how Copilot in Microsoft Teams knows when to execute this particular plugin that I've built in this case, when someone asked about supplier metrics and performance.

But of course, I want to take actions and do things in response to that question. If I click down here, I can use something that we call Generative Answers. Generative Answers allow me to register websites, SharePoint sites and more, which are then synthesized and replied back to the user to answer their question. What's great is I can hard code exactly the data sources I want.

If I go click through and use this Generative Answers, I'm going to get a great experience where I can configure up to four different websites and four different SharePoint sites, where I'll go pull information from to answer the question. This is just as easy as putting in the URL or selecting the necessary websites and documents. It's not just web or SharePoint, I'm also able to use Azure AI Studio and other Azure AI capabilities to extend my Copilot. So, I can start from low code and go all the way to pro code.

For this case, I already have a SharePoint site where I have documents about supplier information, so I'll just paste in that URL and add it, and this is how I can guide exactly the reply I want inside of my plugin. To show what it's using, I'll just click through this URL and look at that supplier factsheet. It's 20 pages, unstructured data, tables and information about all the different suppliers that I use inside of my enterprise.

This is a great way to go make sure that all the different supply associates like me can understand exactly what's happening when I'm using Copilot in Teams. I switch back

over to Copilot Studio, select the input text from the question the user asked to pass it into that data source, and I'm able to get all of this wired up very easily and very quickly.

Now that I have that, I just go up and save and publish. What's great is when I publish this conversational plugin, it's not just for me. It can also be for people on my team, or if I'm in the IT organization, I can publish it for the entire company. So I switch back over here, ask that same question I did before, "Where are the lead times with supplier Alpine Ski house?" and I get back an answer. This answer is from that document. Within 10 days of order placement, that's the average delivery time.

This is great, it's good to get started, but the reality is I want to be more interactive and use the data that's stored in my systems of record. My back office knows exactly the expectations for the suppliers. So, I'm going to go back to Copilot Studio and I'm going to create a new plugin.

For this plugin, I am going to use a different set of data sources. The first thing I'll do is I'll have the same trigger as before, so I'll paste that in because it's the same types of questions that users are going to ask, but I'm going to use something different. I'm going to call to my SAP instance and a customer REST API defined, to bind together this information and give an answer that's exactly right at that point in time.

For step one, I'm going to go down to that call in action, and here we have those 1,100-plus prebuilt connectors and plugins. Whether it's SAP ServiceNow, Workday and more, you can connect to it seamlessly in Copilot Studio. I'll just search for SAP, and I'll be able to get back all the information I care about for the different actions I can execute against my SAP instance. In this case, I'm just going to read from an SAP table, and I'm just going to go configure a few other things like the table name that I care about and more.

This will allow me to easily pull from that information, no matter where it's hosted, using the on-premise data gateway if needed to connect on premise, or the standard connectors you already know from Power Platform. But that's only step one.

I then want to take the result from SAP, and I want to pass it to my custom REST endpoint. And I can do that inside of Copilot Studio by calling this action, and just like we have the 1,100 prebuilt connectors and plugins, you can add as the 1,100 and first by registering your own backend systems.

In this case, the order delivery rates' API is something that I've created and configured, hosted inside of Azure. And of course, I like Azure API Management, so that's where I registered my endpoints as well. So I click through that, and what I start to do is I'm sequencing a bunch of individual API calls to fulfill this question from the user right inside of Copilot. This sequencing is important because I want to control exactly how I pass information between these two systems to go generate my response.

In here, what I've done is I've highlighted something called the Generative Response. The reality is I don't want to respond back with JSON or unformatted text. I want to synthesize information from both SAP and my REST API back to the user. I use that Generative Response, I provide the information from the previous steps just by selecting it inside the designer, and then I'm able to provide a prompt and instructions to guide how I want to format the information and the information I particularly want to highlight. All of this comes back in a really easy, simple workflow going across lots of different systems to give exactly that real-time information.

Now, I click back out, and I'm going to do the same thing I did before where I go save and publish to Microsoft Copilot, and this will allow me to go ask that question in real time.

Switching back to Teams and inside of Teams, I'll ask Copilot "What are the lead times with supplier Alpine Ski House?" I get an answer, which orchestrates across those two different steps and tells me it gets fulfilled within 10 days of order placement and has an 86% on-time delivery rate over the last six months. All of this can be wired up inside of Copilot Studio, and you can start with low code or go all the way to pro code as needed.

The last thing I want to talk about real quick is just those connectors and plugins, because this is what we kept hearing from all of our customers, is one of the biggest challenges of Copilot is making it work over your enterprise data. Copilot Studio makes it easy so you can go build and publish extensions to the Microsoft Copilot. And if you want to use Copilot for your own app and your employees, you can build it in Copilot Studio, too. Thank you so much for the time today. I'm going to turn it back over to Scott.

(Applause)

SCOTT GUTHRIE: Thanks, Charles. Our developer tools are the most widely used in the world and as you've seen, we're differentiating them even further with AI. Every organization is going to be able to use them to build transformational AI solutions of their own.

Now, one of the things that makes these tools even more powerful is their native integration with the Microsoft Cloud, and specifically within the Azure application services. Azure application services include a rich set of cloud platform services. You can use these services with any programming language and technology stack to build enterprise-grade, secure and trusted cloud solutions.

A great example of a customer doing this today is Lego, who wanted to modernize their approach to application development to deliver even more innovative experiences. Let's watch a video about how they're doing this using the Azure Kubernetes Service.

VIDEO SEGMENT: The Lego House is this ultimate Lego experience, as I would call this the home of the brick. We believe that you can learn a lot through play. That's also why we have around 10 interactive experiences. You can drive with robots, you can pull

the stop motion video, we have a virtual fish tank. One of the things that we struggled with was that everything was run on-premise, meaning that it's super hard to error-correct, it's super hard to update. The plan is that we will actually rebuild the full house to run on Azure and Azure Kubernetes Services to really create that flexibility and scalability in our digital textile.

Azure empowers our developers in the way that we can utilize a lot more of the technologies built in the whole Lego ecosystem. Suddenly, we can talk together in a different way and explore different opportunities, and maybe hand some of the great things we build here to other parts of the ecosystem whether that's LegoLands or retail stores. The short-term impact will be to make the house even more stable than it is today, providing higher uptime.

Then over time, my dream is that we can basically add new features, try things out, provide new services for our guests' new experiences. We also need to learn on an ongoing basis, and the only way we learn is to bring the technology into the hands of our guests and see if that helps them getting a better learn-from-play experience, and then we can make the house better and better.

(Applause)

SCOTT GUTHRIE: I grew up as a kid playing with Legos and I'll tell you, that video gives me some joy. This week we're releasing some great, new Azure Kubernetes capabilities. Now generally available, the Azure Kubernetes Fleet Manager simplifies governance and compliance of your AKS workloads by providing centralized controls, standardized policies, and automated resource management across multiple Kubernetes clusters.

You can now group multiple AKS clusters together to control access and stage updates in a safe and predictable way, so developers are assured of consistency and compliance with organizational policies.

We've also added several new capabilities to help you manage your costs with optimized resource usage, fine-tune scaling, and more granular visibility into your cost consumption. With our new AI Tool Chain Operator, we're making it easy for Kubernetes customers to self-host and run specialized AI workloads like large language models on top of their AKS clusters.

I've covered a little bit about the Azure application platform. Now what I'd like to do is switch gears and talk about the data platform. To do that, I'm also going to invite Arun Ulagaratchagan, who leads our Azure Data Product Group, on stage to share some of the exciting innovations that we're bringing to the data stack. Please welcome Arun.

(Applause)

ARUN ULAGARATCHAGAN: Thank you so much, Scott. I'm so excited to be here. We all recognize that AI is rapidly changing the world, but data is the fuel that powers AI. Getting enterprise data ready for AI has become a critical priority for every organization. Now, there's good news and bad news. The good news is that there's been a lot of innovation in the data and AI space. The bad news is there's been a lot of innovation in the data and AI space, because it's created a ton of fragmentation of the modern data and AI stack.

This is the 2023 data and AI landscape slide, and every tiny icon on the slide is a product of technology in data and AI. This is the complexity that's facing you today, because the burden is on you to figure out which products to use, and how to bring them together to create business value.

When I talk to customers, the message I consistently hear is this. "Please unify, I want to be the Chief Information Officer, not the Chief Integration Officer. Help me unleash AI on my data estate." This is exactly where Microsoft can help. With the Microsoft Intelligent Data Platform, we have a comprehensive portfolio of products across databases, analytics, AI and governance. And by working with Microsoft, you not only get really strong capabilities in each area, but even more importantly, you can lean on Microsoft to make everything work together seamlessly so your developers can focus on moving your business forward.

Each of these Azure services are designed to easily integrate with the Microsoft Cloud, including support for Microsoft 365, Teams, Dynamics, and of course, the Microsoft Copilot. We have a number of exciting announcements today and let's start with databases.

In the era of AI, there are many demands on modern cloud-native databases. Data is both structured and unstructured. Traffic is globally distributed, and traffic patterns change all the time. Your AI models need to work with your data in your database in real time. That's why we have the most comprehensive portfolio of database products in Azure, offering relational, non-relational open-source and caching solutions.

We also have really deep partnerships with both Oracle and MongoDB, who provide their databases as managed offerings in Azure as well. Now, let's go a little bit deeper, starting with Cosmos DB. Cosmos DB is quickly becoming the database of choice for the world's AI workloads. AI scenarios like the retail shopping assistants need to respond quickly to a large user base in real time, and Cosmos DB's ability to scale and provide single-digit millisecond performance is critical for a great user experience. Cosmos DB's active capabilities can also be used to run the most mission-critical workloads and guarantee five nines availability.

Now, let me give you an example of an application that I'm pretty sure everybody in the audience knows really well, ChatGPT. ChatGPT is the fastest growing consumer product in history, and OpenAI uses Azure Cosmos DB to power ChatGPT's chat conversation history, context and moderation. Cosmos DB executes many billions of transactions

every single day just to keep up with ChatGPT's growth. ChatGPT also runs on top of Azure Kubernetes Service and was developed using GitHub. We're excited this week to announce several new updates to Azure Operational Database Services. Our Azure Cosmos DB vector search is now generally available, and Vector Search enables efficient queries over the semantic meaning of database records. It is often used to implement the RAG pattern retrieval augmented generation to help improve the quality of responses for GenAI by scaling the amount of contextual information that can be brought into the prompt. KPMG recently developed KymChat, an AI agent to assist employees with operational tasks such as research, drafting proposals and communications. KPMG built KymChat using Azure Cosmos DB, and within a short period it was quickly adopted by over 7000 employees and resulted in a 50% productivity increase for these tasks.

Now, let's shift to PostgreSQL. We are pleased to announce two great new capabilities. The first is new Azure extensions that invokes Azure OpenAI to create vector embeddings natively. This makes the experience of building intelligent applications seamless on Azure PostgreSQL. Second, we're excited to introduce a new performance tier for PostgreSQL based upon SSD v 2, which provides five times more transactional throughput at about half the latency.

Well, let's move on to analytics and AI. In May this year, we announced Microsoft Fabric, our new data and analytics solution built from the ground up for the era of AI. Fabric is truly unified at the most fundamental level with unified compute and storage, unified experience, unified governance and a unified business model. This unification dramatically accelerates your time to value and helps you save a ton of cost. We are achieving this unification by moving from one component to a unified stack, from a single database or a data warehouse to all of your data, from GenAI needing to be manually widened to GenAI simply being built in.

Fabric brings together proven technologies like Azure Data Factory, Azure Synapse Analytics and Power BI to help customers go from data to insights to action all within the same unified SaaS product. Microsoft Fabric has seven core workloads, purpose-built for specific personas like a data scientist, a data engineer, a data warehousing professional, a business analyst, and specific tasks.

Importantly, Fabric also unifies the business model across all of these analytics workloads. You just buy one thing, Fabric capacity, and overnight, the capacity might power your data engineering tasks, maybe some data science, maybe some GenAI. And in the morning, as people walk into the office, the capacity moves to Power BI and to SQL. This unified business model helps you substantially reduce your costs. Just like Word, PowerPoint, Excel natively store their data on OneDrive, these core Fabric workloads natively store their data in OneLake. OneLake is Fabric's SaaS, multi-cloud database that is just built in and automatically available to every Fabric tenant. It natively embraces open-source data formats, with all of the data sitting in Apache Parquet and Delta Lake. OneLake means that any customer can connect to any data source into OneLake, so that they can use Microsoft Fabric as a single analytic solution to manage their entire data estate. OneLake not only provides storage out of the box, but also allows

you to virtualize storage wherever it exists on Azure Data Lake Gen2, Amazon S3, and soon Google Storage as well.

Fabric's vision has really resonated with customers. The momentum we have seen in the public preview over the last five months has just been stunning. We have over 25,000 organizations using Fabric today, including two-thirds of the Fortune 500. We're also seeing that the end-to-end value proposition of Fabric is really resonating with customers, with 84% of customers using three plus workloads in Fabric. There are three exciting announcements for Microsoft Fabric here at Ignite, and the first is that Microsoft Fabric is now generally available here at Ignite and is ready for production workloads. Thank you. Even as we make Microsoft Fabric generally available, we're continuing to deliver innovations at a furious pace. We're delighted to announce this week that Copilot in Microsoft Fabric is now in public preview. With Copilot in Fabric, you can help with the entire data and analytics journey, whether it's with data preparation, data engineering, building machine learning models, or getting insights with Power BI.

The third announcement we're making today is a brand-new capability in Fabric called mirroring, with support for Snowflake, Azure SQL DB and Cosmos DB, with more databases coming soon. Mirroring allows customers to work with existing databases as if they're already native to Fabric. For example, you can simply point Fabric at Snowflake or Azure SQL DB, and we create a mirrored, near real-time instance of that database in OneLake, with the data stored in the native format of OneLake, which is Apache Parquet and Delta Lake.

Mirroring enables blazing fast query performance with Power BI's new Direct Lake mode, and it substantially reduces costs because customers don't need to pay for the SQL queries going to Snowflake or Azure SQL DB as an example. But it's not just Power BI, all of the Fabric workloads Spark, AI, GenAI, real-time analytics, everything lights up on top of the metadata instantly. Now, let me show you a quick demo of how all of this comes together.

VIDEO SEGMENT: Let's see how Microsoft Fabric can bring together data from multiple clouds into a unified data estate. We're going to start by creating a new lakehouse in Fabric for Contoso Outdoors, to seamlessly bring together all of their data. All I have to do is give it a name, press, create, and in a couple of seconds I have an enterprise-ready lakehouse ready to go. Now, we can bring data into the lakehouse by using any of the 150 connectors that we have here, but we don't want to do that in this case because we don't want to manually copy the data. Instead, we're going to use Fabric's shortcuts and mirroring capabilities to access all of our data through Fabric with zero ETL.

If we look at where Contoso Outdoors is storing and managing their data today, you can see that they, like a lot of organizations, have data everywhere. In Snowflake, there is the customer loyalty program, Azure Databricks has the sales data, Cosmos DB has the retail order tracking. Dataverse has the customer support systems, and we can get our product inventory and our product catalog from Amazon S3 buckets. With Fabric, we're going to

make it so we can access all of these sources in a single unified lakehouse that is always instantly in sync with the underlying systems with zero ETL.

Let's start with the Amazon S3 bucket, Azure Databricks and Dataverse. For these sources, we can create shortcuts directly to the data since Fabric natively works with data in Delta Parquet format. Let's set up the shortcut to the Amazon S3 bucket. Now I'm going to jump ahead to where we've already set up shortcuts to our sales data from Azure Databricks and our customer support data from Dataverse. We've done shortcuts to more than just tables of data.

Here you can see we've also created shortcuts to images from our product catalog and text documentation files. We can use these for building solutions in Fabric or for other services like Azure AI Studio, to be able to access the data through OneLake. Now, let's tackle our data in Snowflake and Cosmos DB. For these sources we're going to use Fabric's new database mirroring capability. Database mirroring automatically reflects data from Snowflake directly into OneLake and keeps it in sync with every change.

This means my customer loyalty program data from Snowflake can be accessed seamlessly in Fabric with zero ETL, and it's always kept in sync automatically by Fabric. Once the mirroring is up and running, I can switch over to the SQL endpoint and Fabric and my data is ready to be queried. But let's go back to the lakehouse, because not only am I going to query the Snowflake data, but I want to run queries across all of our data at the same time. You can see we've also set up mirroring for Cosmos DB, and I've added it and the Snowflake data here, so now I have data from five different clouds integrated in minutes into a single unified lakehouse in Fabric.

With one click, I can switch to the SQL endpoint in Fabric for the lakehouse and write a single SQL query that joins data from each of these five different clouds. I can now work with all of these sources through Fabric in a unified way. Most importantly, every workload in Fabric can work with the same one copy of my data. If we switch over to the lineage view, you can see this even more clearly. Here's our Snowflake, Azure Databricks, Dataverse, Amazon S3 and Cosmos DB data all coming together with a full view of downstream and upstream lineage ready for my organization to use.

For example, here we've created a machine learning model to predict demand and inventory levels for the upcoming holiday shopping season. This model requires data from all of Contoso Outdoors' different cloud data sources, and it would have taken days or weeks to pull it all together and integrate it. But now, with Fabric, we've cut this down to minutes and integrated the engines and tooling to create the solution right here. Finally, my lakehouse in Fabric has a semantic model automatically created that makes it easy to access the data through BI tools like Power BI. Let's create a new Power BI report and use Copilot to get started.

Let's ask to create a report showing order status for our loyalty program members. Again, this is pulling data from multiple cloud sources. Just in a couple of seconds, Copilot has looked at my data and generated a beautiful report. I can slice and dice, and right away, I

see the low review scores have higher delays, and I can use the report as a starting point to customize it for my own needs. Let's add one more visual to the report. This is a Copilot-powered summary visual that will create a narrative about the data in the report. Let's ask for five bullets with facts about the order status.

Copilot provides a dynamic list of insights that will update with the data, and it provides citations, so I can see where the insights come from. As you've seen, Microsoft Fabric is a game changer for enabling organizations to create a unified data estate, spanning all of their data across clouds, and supporting every analytics workload in one easy to use experience.

ARUN ULUGARATCHAGAN: We're really excited about this because we recognize that you have data everywhere. With Fabric, we simply want to embrace data where it lives and really deliver business value very, very quickly. Now, let's talk about governance. Governance and security are critical to any data estate. Purview is Microsoft's built-in, multi-cloud hybrid solution for data governance. With the GA of Fabric, we are pleased to announce end-to-end governance and security powered by Microsoft Purview. For example, Fabric includes a built-in data catalog, the ability to classify data by different data domains, certify data artifact, C lineage, and apply Microsoft Information Protection labels, which automatically travels with your data even if the data leaves Fabric and goes into Office tools like Excel and PowerPoint. It's exciting, yes.

We're delighted to announce the public preview of the Purview Hub that's just built into Fabric that gives you end-to-end visibility of everything going on in Fabric in one central place. What you've seen so far is the breadth and depth of the capabilities that are available in Azure across the Microsoft Intelligent Data Platform, going across databases, analytics and governance to help you accelerate your transition to the era of AI. Today, we're also thrilled to announce that we have many of the same partners for the Microsoft Intelligent Data Platform, ISVs, that are extending their partnership with us to include deep integration with Microsoft Fabric.

One of the partners we've been working very closely with is Informatica. Informatica is all in on Microsoft Fabric, and we're excited to announce today that Informatica will be delivering its AI powered cloud data management solution as a unified experience in Microsoft Fabric. One of our most strategic partnerships is with Databricks. We are proud of the co-innovation that we've done with Databricks. Azure Databricks and Microsoft Fabric are highly interoperable, as both products are based on a common Open-lakehouse architecture. Now, I'd like to welcome Scott back on stage, and with him, Ali Ghodsi, co-founder and CEO of Databricks, to share more about the incredible work Microsoft and Databricks are doing together. Welcome.

ALI GHODSI: Thank you, Arun.

ARUN ULUGARATCHAGAN: Thank you, Ali, for being here. We're very excited about our partnership and all the work we're doing together to help our customers accelerate their data and AI initiatives.

ALI GHODSI: Thank you so much, Arun. I'm super excited to be here. It's so awesome because now we have the same foundation for Fabric and Azure Databricks. It's all built on this open-source project called Delta Lake, so it'll be really seamless to go between these two projects.

ARUN ULUGARATCHAGAN: Awesome. I think you have some exciting stuff to show us.

ALI GHODSI: Yeah. All right. Very excited.

ARUN ULUGARATCHAGAN: Let's take a look at it.

ALI GHODSI: All right. Basically, Databricks is a unified platform where you have one copy of the data in Delta, which is now shared, and one place to do governance in Unity catalog, so let's start here. You can see here there's a SQL data warehouse here, there is data engineering DTL for delta and machine learning but let me go over to the catalog here. This is the Unity Catalog; inside Unity Catalog you have all your tables. You can actually even federate into other systems, so we have a table that we've federated from Snowflake here, you can federate BigQuery or other systems.

But what I want to really do is go look at this one table that we have in the system, which is a churn table, so churn features. You can see here if we click on it that we have different columns in the table, we have some sample data that's sitting in here, there's permissions. You can set up attribute-based access control, you can look at the lineage so you can see what the upstream downstream tables are, you can even see the whole lineage graph here. What we can do now is giving all this semantic information about lineage metadata, the AI has data intelligence to actually generate the text for what this table does. We can actually click here, and for each of the columns, we can see that it's describing in English what they each do. We have an English description. This now enables us to do semantic searches. So we can go to search, we can type in English, "find information about platform churn." And it now will actually leverage that English-generated data intelligence and give us the table that's important for this. This is super critical because it's really hard in organizations to find the right data.

Let's look at here in SQL, I'm going to actually do some queries here. "Show me the number of churned users for premium platform in the past fiscal year." It's going to generate SQL for us here, but now it actually understood the data intelligence for these words, "premium," "platform" or "fiscal year." Let's run this query. Right, so we've got the results. Actually, it generated a pretty good SQL for us, so now, what we're going to do is use this to do some more intelligence with this. Super exciting.

SCOTT GUTHRIE: It's awesome to see how you're using AI, and specifically, the Azure OpenAI Service, directly in the experience. That's awesome.

ALI GHODSI: Let me show you what we can do actually with Azure OpenAI and do some build some RAG chatbot together. Let's start here. What I'm going to do is I'm going to go here, and I'm going to upload a bunch of PDFs on documentation in Databricks here to Unity catalog volume. It now takes each of these and it actually creates delta tables from these automatically. If we go click here on Databricks documentation, we can see that now it has the sample data uploaded, it's actually chunked it, it understands the contents.

We can actually now say: Let's create a vector search database with this. We'll give it a name, and we have to then pick a primary key so that the vector search can use, and we use an endpoint, and then we use an embedding model. I'm actually just going to use Azure OpenAI data model here, and I'm going to say that continuously update this vector database. We can click here; we can see that now the vector database is up and running. What we're going to do next is create an endpoint. So we're going to serve this model, let's create an LLM using Azure OpenAI chatbots and create the serving endpoints. So it's off now, and it's creating an LLM endpoint that we can call into.

I prepared a notebook here that we can see. This notebook has a chatbot in it, and we can actually ask a question that calls that endpoint now. "How can I publish my unity catalog scheme to Power BI?" You can see this hits the endpoint that goes to the vector search and it answers it. This is awesome. This now also gets automatically stored in a delta table. So we can see here all the requests, all the monitoring, and we can easily just push this now to Power BI. It pushes the dataset into Power BI. We can go here and say, "Automatically create a report based on the delta table that we have." We can see here that we can see the toxicity or all the responses that got recorded, so it's really, really tightly integrated now into Power BI. You can see here with a few clicks we were able to build our own chatbot, leverage Azure OpenAI, and get the insights published into Power BI.

SCOTT GUTHRIE: That's fantastic. It's awesome to see how easy it is to use Azure Databricks to create a modern GenAI application on Azure.

ALI GHODSI: We're super, super excited. We're super excited about doubling down on the partnership and seeing what we can do together to integrate this even more and more with Fabric.

SCOTT GUTHRIE: Thank you, Ali.

ALI GHODSI: Thank you.

ARUN ULUGARATCHAGAN: Thank you so much, Ali.

SCOTT GUTHRIE: So data is the fuel that powers AI, and the new capabilities that we're releasing this week are really going to enable you to unify and simplify data management projects and have a data estate that's ready to build awesome AI solutions. Now that we've covered the data stack, let's now talk about our Azure AI platform and how this all comes together. Now, the pace of AI innovation at Microsoft has been pretty intense over the last year, and the reason we've been able to move at such a fast pace is because we've built all of our Copilots on top of one platform, which is Azure AI. This is the same AI platform that you can use to build copilots of your own.

Azure AI includes several categories of AI capabilities, including our Advanced AI Services, our purpose-built AI search Azure Machine Learning, and our Azure AI studio. This week, we're excited to announce the previews of GPT4 Turbo, Dall-E 3 and GPT4 Turbo with Vision. GPT4 Turbo offers lower pricing, extended prompt length, and structured JSON formatting for improved efficiency and control. We're also opening a new chapter to GenAI, which is multi-modal capabilities that are going to enable you to build generative AI applications that go beyond purely text-based scenarios.

Developers can now enable apps to literally see and perform object detection, as well as enable apps to understand and make inferences such as video analysis from visual input prompts. The combination of GPT4 Turbo with Vision, integrated with the rest of our Azure AI Vision capabilities, we're really going to enable you to build the next generation of interactive AI applications. Now, with Azure AI, you can use any GenAI model, including obviously, the Azure OpenAI models, open-source models as well, and models provided by other commercial companies. We're excited today to announce our new Azure Model-As-A-Service offering that's built into Azure AI.

Model-As-A-Service enables you to use foundational models in your applications without having to manage the underlying GPU infrastructure. This includes Llama2 from Meta, Mistral's premium models, G42's Jais models, and Command from Cohere. Again, this eliminates complexity and enables you to use the best models for any particular task. What makes our Azure AI Service so powerful is that they enable you to also ground and fine-tune these AI models using your own data. This enables you to get much more intelligent and tailored outputs from these AI models. We're making it easy to do this with both Microsoft Fabric as well as all the other Azure Data Services. And it's really important when you do this, when you ground your models with your data, that you trust your AI cloud provider. Microsoft runs on trust, which means Azure AI runs on trust. And Microsoft's committed to security, privacy and compliance across everything we do. And our approach to AI is no different.

When you use Azure AI, your use of the service is isolated from every other customer. Your data is not used to train the foundational models used by others. You don't have to worry about anyone other than your organization benefiting from AI that's based on your data. And your data and AI models are protected at every step by the most comprehensive enterprise compliance and security controls in the industry.

In September, Microsoft announced the Copilot Copyright Commitment, which is a new benefit to defend all commercial customers from copyright claims related to their use of the Microsoft Copilot product. And today, we're excited to announce that Microsoft is also broadening this commitment to include the Azure OpenAI service as well. This new benefit is called the Customer Copyright Commitment, and it's going to help you look to accelerate your use of generative AI beyond Microsoft Copilot and be able to use it inside your own custom applications as well.

So it's been an incredible year in terms of AI information, and to share more about some of the great Azure AI enhancements coming out this week, I'd like to invite Eric Boyd, who leads the Azure Engineering team, to the stage.

Please welcome Eric.

ERIC BOYD: Thank you, Scott. Customers have been so excited to build generative AI applications, and they tell us that there are a lot of different components that they need to bring together to build these applications. That's why we built the Azure AI Studio, the one place with all of the tools you need to build rich and powerful generative AI applications. And of course, that all starts with data. You need data from your enterprise to personalize the application to your specific needs. And you can bring that data from wherever it happens to be including connectors to Cosmos DB and Microsoft Fabric. And then you bring that data into a search and retrieval system to make sure you're feeding the large language model with the best data that you have possible for each individualized query.

And of course, you can choose the large language model that you want directly from our model catalog. You can use the latest foundation models from OpenAI, the most powerful models out there. You can also use the greatest open-source models as well, directly in that same model catalog. You then manage the full lifecycle of building this application with Prompt Flow tool that I'm happy to announce is now generally available. And of course, underlying all of this is the Content Safety System to make sure that the application you build will be safe and secure.

Let's talk a little bit more about that Vector Search engine. The search engine is one of the most important parts of the AI application stack, making sure that with each query that you get, you're giving the large language model the most relevant data on what you're trying to answer.

Vector Search is the new technology that helps match the best semantic meaning of what you're asking for with the best relevant data in the data set that you have. Azure AI Search is the preferred platform for companies building generative AI models, for those who need the performance, relevance and scale.

In fact, you may have heard Satya mention yesterday that OpenAI, at their Developer Day conference, last week announced a number of generative AI RAG features. All of

those features are powered by Azure AI Search, because they wanted the relevance and the performance and scale that comes with this platform.

Azure AI Search is not just the Vector Search engine. We bring all of the relevance, to get the semantic meaning relevance and semantic matching from vector searching, and we combine it with all of the techniques that we've spent decades developing, building highly relevant search engines like Bing to bring things like keyword matching, misspelled words, synonyms and all of that into a single corpus, and then we re-rank that to make sure we're providing the most semantically relevant data for each and every query that you're doing. And I'm excited to announce that Vector Search is also generally available in Azure AI Search.

Now, as you're building your generative AI applications, what we just talked about is often called retrieval augmented generation or RAG, and that's generally the most powerful way to build these applications, but you can take things one step further. What fine-tuning does is it enables you to actually take the data that you're using and customize and personalize the innards of the large language model itself with your own customers data.

And I'm happy to announce that we're enabling fine-tuning on both GPT-3.5 Turbo as well as on GPT-4, and we'll bring all of those capabilities of these large language models and generative AI applications into the single place the Azure AI Studio. But more than talking about it, I'd love to show you just how easy it is. And so I'd like to invite Seth Juarez, a principal programming manager on the Azure AI team, to come out and walk us through this.

Seth?

SETH JUAREZ: How's it going, Eric?

ERIC BOYD: It's going great. You're looking great, Seth. Nice shoes by the way.

SETH JUAREZ: Thank you, yeah, I love these shoes.

I'm pretty excited. I work in the Azure AI engineering group, but I really split my time with Contoso Outdoors, and so that's really where I do all my demos. I'm going to show you three things, now that I work on the Contoso Outdoors demos, as to how AI can help you in three ways.

No. 1, content. No. 2, support. And No. 3, sales. I'm going to show you how Azure AI can help you with all of those things. What you are looking at here is the Contoso Outdoors company website. It's actually quite beautiful. When you look at it, you're also going to see some wonderful text here. I am not this creative, and so GPT-4 Turbo made this for me. That's the first thing.

The second thing I want to show you is this wonderful Trailmaster X4 tent. Those beautiful pictures were created by DALL-E 3 because I'm not artistic. And you can see here these are wonderful tents. And so notice that, with these new Azure generative models, you're able to actually build amazing content to delight your customers as they show up. That's the first thing content.

No. 2, support. Even though I made this beautiful website that has everything customers will need, for example, to know what to do with their tents, we all know here, in this very room, that every single one of those customers will not read this, and they're going to call us instead, right?

So how can we solve this problem? It turns out that the support team said, "Hey, why not make a chat that answers the exact questions that our customers ask based on the exact facts that belong to our product?"

So I'm going to go here, and I'm going to ask a question, right here. And the question is, "How should I take care of my tent?" And notice that in a matter of like, three seconds – two, one, you get an answer, and not only is that answer a good answer, but it also has links to basically this page that has the information in it.

So how do we accomplish this? It's pretty cool, right? And notice, it's in your website. It's in the beautiful website that I created for you. So how do we do this?

Well, to do that, we're going to use the brand new experience called Azure AI Studio, and I'm just going to show you a couple of things in here. The first is, when you build things now, you build things inside of projects, which is pretty much what we do.

So let's go to the Contoso Store, and let me show you something here, because I want to explain what actually is happening when you add your data to these models. Most people think that we're pushing data into the model, but we are not. What we're doing is we're creating a search engine, and when I say "we," I mean like in the Packers, "We won the game," but what we're doing is we're actually creating a search index in Azure AI Search, and the source of that data really can be anywhere.

So let me show you. If I go over here, and I go to indexes, and I create a new index, notice that I can actually connect to something like Microsoft OneLake, where actually, like where actually all the product manuals live, you saw that earlier. And what it does is it takes those manuals, it chunks that data and puts it into Azure AI Search, so that when your customers ask a question, that question then is searched, and the information is put into the prompt and then sent to the large language model. That's what's actually happened.

Let me show you. So I've already set this up. You saw there that we already have the Contoso Manual Index, and so if I go over here, I've hooked it up. And so when I asked the question again, inside of this experience, as you test it, you're going to notice that in about three seconds, you're going to get the answer, and I want you to notice something.

It's almost exactly the same as the answer we got here. I think it's almost exactly the same thing. And the reason why is because we give you the code, the exact code to do this, and you can also export this to a problem. Pretty cool, right? And it's something you can do today. That's right.

The third thing is sales. If you recall, Eric mentioned my shoes. I think they're the coolest shoes ever, and I know Eric liked them, and so what I'm going to do is, I'm going to see if I can't get him some of these shoes here at the Contoso Outdoors products company here.

So let's see, my friend Eric really likes these. I'm not shouting – oh, my goodness, let me – I'm like shouting at the computer, and I don't want it to get mad at me.

So what I'm going to do is I'm going to open this up, and I'm going to put my shoe in here, because he likes it. And I'm going to ask you to question, because the reality is that, when we're looking at the real world, it turns out that the real world is actually multimodal, right?

So when I give it the shoe, and I asked it a question, I want it to answer, based upon what I just gave it, and you're going to see something here really cool. "It's great to hear that your friend Eric has good taste." That's debatable. "Based upon your request for a pair of gray outdoor sandals in red and gray," that's pretty specific, and it's suggesting sandals from our company.

That's pretty cool, right? Let me show you how this is done. This happens inside Azure AI Studio, and I just want to show you a couple of things. The first is that it uses Azure AI Studio's feature called Prompt Flow. And the first thing I want to show you is that, with GPT-4 Turbo with Vision, you can actually put images into prompt flow, which is pretty exciting. So that's the first thing.

The second thing I want to show you is the actual prompt, because you're probably going, "Well, how do you put an image in there?" Well, it turns out that the prompt is just like you would expect, text. And then here is the image that you put in. It's pretty cool, right?

And then the last thing I want to show you is that there is a really powerful extensibility bundle with Prompt Flow that allows you to have custom tools built directly into the UI, so things like Langchain, semantic kernel, all of those things can live right inside of your UI.

And the last thing I want to show you when we finish up is that you can deploy these things with one click. For example, you can see how many times I did this. If you go to the Contoso Retail Store #10, you can actually switch between different deployments live so that if you have a new prompt flow that comes out because you fixed something, you can have that available to your customers immediately.

You've already started building these things. I looked at our telemetry last night. I'm excited to see what you can build with this, go to ai.azure.com, and I'm excited to see what kind of experience you will enable.

Thank you so much.

ERIC BOYD: Thanks so much, Seth. I have been led to believe that keeping your shoes on was required at these keynotes, but Seth has proven me wrong. We've had so much interest from customers really coming to the Azure OpenAI service, because they know that they can get the latest models from OpenAI directly with the security and privacy of Azure built right in.

We've had over 18,000 customers put their trust in Azure OpenAI's service, and built just such stunning applications on top of it. Let's take a look at what one of these customers, Siemens, has been doing.

VIDEO SEGMENT: At Siemens Software, we're focused on bringing together engineering and manufacturing all the way down to the shop floor in a unified experience. Imagine you're on a factory line and you see a problem. Maybe you have gloves on. Maybe you're not speaking a language that is the same as the software that has been implemented to track that factory line. The reality is problems won't be reported if it's too hard. AI allows problems to be reported in native language and automatically translated to the common language. At Siemens, we are all in on AI, and Azure AI Studio is a key tool for us.

We chose Azure AI for its comprehensive suite of tools that streamline our workflows from prototyping to deployment and production of services, from your first approach, accelerating our adoption of advanced machine learning technologies. We have utilized Azure AI's robust infrastructure to deliver a secure and highly available team in the Teams app. Azure AI's network isolation and SLA-backed availability were key in meeting our enterprise data objectives.

With the Tell Me feature, we have now enabled the field engineers to be able to speak in their own language to report issues and create data items in the Teams center. We believe that now the field engineers will be able to feel more included in the workplace. The managers or the engineers responsible for resolving the issue can take necessary action steps.

We are anticipating more problems reported, better problems being reported and faster resolution times. And using Azure AI allows us to do that.

ERIC BOYD: Now, when you're building generative AI applications, safety is not an optional feature, it's a requirement. Safety is something that we've thought very deeply about at Microsoft and work to really make sure we're living up to the responsible AI principles that we've espoused. And I'm happy to announce that the Azure AI Content

Safety System is generally available. This is the system that we use to protect the Microsoft Copilot, and we're making it available to customers for their own needs.

I'd like to invite Sarah Bird out, who leads the Responsible AI team here at Microsoft, to come and tell us more about the Azure AI Content Safety System.

Sarah?

SARAH BIRD: Thanks, Eric. You saw Seth show you how Contoso can use AI to build a better retail experience. Now, let's look at how we make it safe. We know from our own experiences building Microsoft Copilot that this requires a defense in depth or layered approach to safety. The first two layers are built directly into the Azure AI platform. Starting with the core, we've worked with OpenAI to build safety into the model so that it can recognize problematic queries and respond appropriately.

The next layer is Azure AI Content Safety, an AI-powered safety system designed to support real-time detection of problematic content. Starting today, we have new enhanced features based on our own experiences and feedback from our customers.

Let's go back to AI Studio to take a look. Here in the playground, I can interact with Seth's Contoso application. In recent months, we've seen the rise of new types of attacks on AI systems, such as jailbreak and prompt injection attacks. These attacks are designed to confuse the model and have it break out of its system instructions.

Let's see how the Contoso system responds to one of them. Here I've entered a common one that tries to get the model to frame harmful information as a story. It looks like the text successfully confused the model and it generated harmful content. Fortunately, since this is built on Azure OpenAI, content safety is built in and is already monitoring for problematic inputs and outputs. It caught the issue and filtered the output, preventing anyone from accessing it.

However, not all jailbreaks are attempting to produce harmful content. Here, I'm going to use a DAN style jailbreak to confuse the model, and then ask for some tax advice from Contoso, which is not something that Contoso wants its system to be doing.

Let's see if we can do better. With the new safety capabilities we released today, I'm going to create a custom content filter. Since I don't expect there to be any need for sensitive content in this application, I'm going to set everything to filter starting at low severity to give me maximum safety.

Next, I'm going to enable the new filters to help address emerging risks specific to generative AI applications. The jailbreak filter detects incoming jailbreak attempts, and these filter protected text and code, which means they filter content that may be copyrighted, such as song lyrics that might accidentally appear in the output, so I don't use information that I don't intend to. Now that I've made my selection, I can easily attach the setting to the deployment endpoint to update its behavior.

Let's go back to the playground and see how it's working now. I'm going to retry that same DAN attack again. This time the safety system recognizes that this is a jailbreak attempt and blocks it, protecting Contoso and its users.

As Eric shared, Azure AD content safety is now generally available. We've added more controls for customers to tune settings for their applications. And starting today, the jailbreak and protected material features I demoed are available in public preview.

The platform safety layers provide a great foundation. However, the application layers are equally important for tailoring the safety to the context and the domain. The heart of developing any generative AI application is guiding the model with effective meta prompt instructions and giving it appropriate grounding data.

Today, we are releasing new meta prompt templates, designed and tested for safety, that you can use to get started when building your own meta prompt. For Contoso, Seth has used these, and he's grounding with Contoso product catalog and customer information, and so I think we're in great shape.

However, we've only tested a few examples, which isn't enough to safely ship. So let's go back to AI Studio to properly evaluate it.

In the evaluation tab, I can create a new evaluation run to test Seth's flow for different metrics. However, since I want to test with hundreds of examples, it's going to take a couple minutes to run. So let's look at one that's already completed. Here I've tested the application for groundedness because I want to make sure it's accurately reflecting Contoso's information and relevance and retrieval score, because I want to ensure it is effectively responding to customers. These look great, but we should also check the safety scores for the system. I've added a custom metric to test how well the safety system deflects adversarial attacks.

Let's compare it to a test before the jailbreak filter was turned on. These are highly adversarial tests, so we expect some amount of defects. You can see here that the system, with the jailbreak filter turned on, successfully defends against more attacks, so the defect rate is much lower.

Now that we've tested the system, we can responsibly deploy and bring the power of generative AI to Contoso's shopping experiences.

You've seen how AI Studio is a one-stop shop to confidently build and evaluate a copilot using state-of-the-art responsible AI tools. We're making it easier to build real-world, mission-critical, safe and secure AI solutions that you can trust and use in your business. And all of this is available in Azure AI Studio.

Scott, back to you.

SCOTT GUTHRIE: Thanks, Sarah.

Azure offers a fully integrated stack that is purpose built for AI-powered applications, giving you a competitive edge and superior performance. With pioneering AI technology, a deeply integrated platform and leading cloud security, Azure is your partner for delivering next-generation transformational experiences.

Now that we've covered Azure AI, let's now dive deep into the infrastructure that's powering all of what you've seen today. We've invested heavily, over many years now, to make Azure the place to do cutting-edge AI innovation. Azure is the world's AI supercomputer, and it's the infrastructure that's not just powering ChatGPT and all of the Microsoft copilots and AI products and services from Microsoft, but it's also the AI supercomputer that was used to train the large foundational OpenAI models.

And as you heard Satya talk about yesterday, Microsoft is optimizing and innovating at every layer of the infrastructure stack to deliver unmatched performance and capabilities. And we're doing this from systems to silicon, including the Maia and Cobalt silicon chips that we announced just yesterday. And we're taking the learnings from running the largest and most complex AI and compute workloads in the cloud, including services like Microsoft 365, Teams, ChatGPT, Xbox, Bing and GitHub Copilot, and using that feedback loop to tune and make the best possible infrastructure underneath it.

What this means is that when you use Azure in the Microsoft Cloud, you get industry leading price, performance and feature set capability. Azure is also purpose built to support all workloads, and that includes those spanning on-prem, edge and multi-cloud environments. We know enterprises value the modularity, integration and simplicity that Azure provides, and with Azure Arc you can support running resources and workloads wherever they're needed, regardless of location. The adaptive cloud approach enables organizations to unify previously siloed teams, distributed sites and sprawling systems, and more than 21,000 organizations today are now using the Azure Arc service to extend the Azure platform across their hybrid environments to manage your resources consistently, using Azure tools and services. And built on the capabilities of Azure Arc you can use Azure's adaptive cloud approach to operate from one AI enhanced management layer, scale applications across boundaries and connect edge data from distributed systems.

We're continuing this work this week to expand our adaptive cloud approach delivered through Azure Arc. For VMware customers, we're pleased to announce the general availability of VM vSphere support enabled by Azure Arc. This enables hybrid VMware environments to now run and use Azure services, including on your edge or in your on-premises environments.

And now, also in preview, is our Azure IoT operations, also enabled by Azure Arc. This makes it easier and faster to build IoT solutions that bring you near real-time decisions, all backed by AI-driven insights.

As we've seen, Azure supports workloads anywhere, and that approach expands to purpose built for migrating every enterprise workload. And to share some of the great work we're doing here, I'd like to invite Erin Chapple, who leads our Azure Core Product and Design team, on stage to show off some of the great new enhancements. Please welcome Erin.

(Applause.)

ERIN CHAPPLE: Thanks, Scott. To bring you the best solutions for migrating these enterprise applications, we're partnering deeply with enterprise leaders that you rely on to ensure Azure is purpose built to support your needs for solutions, using SAP, VMware, NetApp, Epic, Citrix and Oracle. And our recent example is a partnership with Oracle.

As we announced in September, Oracle and Microsoft have been working together on a new offering, Oracle Database@Azure. You can now use Oracle database services running on Oracle Exadata hardware, deployed as a service in Azure datacenters. And this streamlines the migration of Oracle workloads to Azure, so you can integrate and innovate with the breadth of Microsoft Cloud services, like Azure AI, and our application platform and developer services.

You'll be able to purchase this fully managed Oracle service from the Azure Marketplace with the same features and pricing as Oracle Database Services on the Oracle Cloud. And you can use your existing Oracle and Azure commitments and licenses to do so.

Let me show an example of how this works.

I can get started using Oracle Database@Azure the same way I do with all my Azure services, just by clicking the link in the portal and starting the process to set up my environment. After choosing my subscription and giving it a name, I can tell Azure where to provision this instance, so it's available in the Azure region and availability zone of my choice. This allows me to ensure the best performance and availability at the very beginning of the migration journey.

Next, I can customize, compute and storage ratios in my environment, according to the workload requirements. This gives me the control to optimize current total cost of ownership and the flexibility to plan for future growth. Similar to other services in Azure, I can specify my preferred maintenance window. This helps me manage interruptions to this very mission critical solution. Last, I review the information and Azure will have Oracle provision this Exadata instance for me right here in the Azure Data Center.

Once provisioned, I can see my Exadata instance ready in Azure. After a simple review of the information to make sure it's correct, I'm ready to create the virtual cluster for my database environment and connect it to the rest of my Azure services. Once I select the Exadata VM cluster, I can confirm the details about the environment. This simple experience ensures the cluster will be set up securely out of the gate.

Once I'm done with the basics, I'm ready to define my application environment. This splendid experience is designed to feel familiar to both existing Azure and OCI customers. Azure allows me to specify the compute and storage I want to allocate to this specific cluster. With these tools, I can optimize application performance and cost while maximizing the benefits of Exadata for this specific Oracle Database@Azure environment.

And finally, the last thing is to specify the Azure v-net used to provide secure connectivity between my Oracle database and the rest of Azure. Microsoft and Oracle have already done the work to physically and securely connect the systems across datacenters with minimal latency. And Azure v-net ensures secure access between all the services in the end-to-end application.

Now that I have a database application running in Azure, let's check in and see how it's doing.

First thing, I look at my applications dashboard in Azure Monitor to check out the health and performance of my system. The out-of-the-box integration of Oracle Database@Azure and Azure Monitor makes it easy for anyone to create custom dashboards, and even create a single pane of glass view across the entire application landscape.

Here, I see the performance metrics of the VM cluster itself, and everything looks healthy. I can then switch to look at another dashboard, containing the metrics and utilization for the database I created.

Here, I see the Oracle database-specific metrics, allowing me to determine if the database is performing optimally. Again, everything looks good.

Thanks to Oracle Database@Azure, I can use my Azure tools to monitor and maintain the health of my systems and have confidence knowing that everything is operating according to plan.

Finally, now that I am sure my system is healthy and running well, I'd like to use Azure AI to help users make decisions about what product to choose.

Earlier, Seth was on stage showing how Contoso Outdoors use AI Studio and Prompt Flow to help customers select the right product. I'm going to extend the prompt flow Seth used to include structured data from the Oracle database.

Here, you see a prompt flow inside Azure AI Studio, using Oracle data as a source for grounding the model with facts from my database. I want to extend the prompt flow from before to include specific customer data to further personalize the results.

Because of the extensibility model, I can create a customized Oracle SQL query tool for the database in Azure. Notice that all the pertinent connection data is managed using

built-in prompt flow connections, and also notice how I can use standard parameterized SQL queries. In this case, I retrieved customer information along with their previous orders to ground the problem.

Now take a look at how the model responds with a more personalized and relevant answer for customers. It knows exactly who the customer is, along with their previous purchases, and gives a very personalized experience when they interact with our website.

To recap, in just a few short minutes, I've shown you how to easily get started migrating your Exadata solutions to Oracle Database@Azure, the power of using Microsoft and Oracle solutions together in the cloud, and how easy it is to add AI to modernize your applications around Oracle data.

(Applause.)

Today, I'm pleased to announce that Oracle Database@Azure will be generally available starting this December. Microsoft is the only hyperscalers other than Oracle to offer Oracle Database services on Exadata infrastructure. To learn more, reach out to your Microsoft or Oracle account team.

Now, we talked about a number of innovation spanning the entirety of Azure today. Taking a broader view of the entirety of your IT environment, it comprises many services with thousands of resources, all different types. And you need to manage all of this. Enter Microsoft Copilot for Azure.

Copilot for Azure will help you gain access to all the knowledge held within your growing ITS states. By combining advances in LLM with a rich data, policy, governance and access control, and the Azure Resource Manager and the Azure Resource Graph, Copilot for Azure lets you easily interact with your entire cloud environment. And it's your agent that helps you act on all of that knowledge. Let me give you a preview.

Copilot for Azure can be accessed anywhere at any time from the Azure portal. And with it, you can talk to Azure naturally, asking questions, like how many resources do I have? Copilot understands and quickly answers your question and invites you to ask more. Related questions are suggested, and I can continue to explore, asking, what are the most common resources? How many VMs do I have? And what regions have the most?

Copilot also gives me the Azure Resource Graph Query I can run to further explore the data and interact between the main query window and the conversation with Copilot, going back to ask what was created in the last five hours.

But operations are more than queries. I can perform management operations as well, such as, tag these VMs with demo, or, show me the memory usage trends to drill into the specific metrics. I can even run management operations, like reboot this machine.

Beyond the resource model, you can use Copilot's knowledge, enriched with Azure documentation and best practices, your billing data, and service logs and metrics to instantly generate deep insights, easily discover new cloud functionality and more quickly get complex tasks done.

Cloud environments have hundreds of rapidly evolving offerings, and understanding their capabilities and how they compose can be a full-time job. When something isn't working right, even the most accomplished expert has a broad surface area to start troubleshooting. Copilot for Azure provides AI-guided troubleshooting that helps you more rapidly identify issues and solutions, so you can solve problems faster.

For example, Copilot is deeply aware of individual services like the Azure Kubernetes Service, providing a huge boost to productivity when troubleshooting complex issues by surfacing the right tools, suggestions and mitigation.

I got an alert that a node has high CPU utilization and can ask Copilot, what should I do? Based on the suggestions offered, I select "how can I add more nodes to my cluster." Copilot then helps me navigate the portal tooling to correct settings to scale, rapidly identifying the solutions so I can finish this troubleshooting quickly.

Navigating the initial learning curve to adopt new technologies presents an opportunity for teams to acquire valuable skills to revitalize an organization. Whether you're starting with a simple web app or choosing to modernize your application to be cloud-native on Azure Kubernetes Service, Copilot for Azure helps you learn, understand, recommend, design, and deploy the right services and infrastructure for your workload.

Copilot always has the most up-to-date data and information so you can design, build and deploy your organization solutions in a more scalable, flexible and cost-effective way. I can ask questions like, what service should I use to run a high scale web application? How does that compare with other services? What data store would I add for a read-heavy, no-SQL workloads? And then Copilot makes it easy to deploy, using CLI or ARM and bicep templates.

Application teams often need to optimize for performance and cost. It's critical for teams to get insights into spending, and get recommendations on how to optimize forecast, or perform what if analysis.

Copilot can help you understand your invoice and spending, changes and outliers. For example, Copilot can summarize my costs for the last six months, understand what subscriptions caused the increase in December, get recommendations for how can we optimize our costs, and get answers based upon the up-to-date cost management usage and billing data. I can also get recommendations for optimizations at the individual service level. With Copilot you can more quickly optimize your cloud environment on the dimensions most important to you.

I'm excited to announce the preview of Microsoft Copilot for Azure, your Copilot for cloud operations and management.

(Applause.)

Now, more than 15,000 Microsoft employees and a handful of private preview customers are already using it today. And we're excited to have you start using it, too. And we look forward to learning together.

With that, back to you, Scott.

(Applause.)

SCOTT GUTHRIE: Thanks so much, Erin.

The Microsoft Cloud is the place for every developer, every IT professional and every organization on the planet to innovate using AI. We're really looking forward to building some great solutions together with you. We have an exciting future ahead of us.

Thank you so much and enjoy the rest of Ignite.

(Applause.)

END