White Paper

Enabling Cloud for All Compute Needs: Azure's Compute Portfolio

Sponsored by: Microsoft

Deepak Mohan Donna Nitchie
July 2020

IDC OPINION

IT leaders have never had so much choice in their infrastructure strategies, both on-premises and in a hybrid or off-premises world. There has been a growing acceptance of public cloud infrastructure as a service (IaaS) among enterprise IT organizations. Cloud is now seen as a necessary component of an organization's IT environment, and cloud computing is being adopted for nearly all workloads and use cases today.

Cloud adoption is no longer in the realm of new or experimental workloads but a critical part of infrastructure modernization for existing "legacy" enterprise applications running on premises. Thus migration and modernization of these applications has become an integral part of the scope of the cloud journey and of the considerations when planning adoption of cloud services. This focus on cloud adoption has accelerated in 2020, owing to the COVID-19-related uncertainty and the desire to operate on a flexible yet cost-effective digital back end.

Through Azure, Microsoft has accelerated its focus and investment on delivering a comprehensive cloud platform for enterprise IT workload needs. Microsoft's position as the second-largest public cloud IaaS provider and the 59% YoY growth for Azure reported in the FY 3Q20 results are a testament to both this investment and the momentum expansion of the Azure platform. Core to this is the Azure compute portfolio and its ability to meet the wide-ranging priorities in the market.

Today, IDC believes Azure compute offers customers the following key value propositions:

- Breadth of compute portfolio covering a broad spectrum of workload needs
- Cost-effectiveness — for both Linux- and Windows OS-based workloads
- Security as a core component — built on decades of enterprise IT experience
- Hybrid capabilities — delivering a unified on-premises/off-premises/edge experience
- Management services for a diverse cloud environment
- Continued momentum of expansion, on Azure adoption enablers and the broader Azure ecosystem of higher-layer services and, critically, on the core compute capabilities — including new processors, leading-edge high-performance computing (HPC) offerings, and new cost optimization options

Recent enhancements and announcements around Azure compute experience for customers include:

- **Azure Arc** – Enabling customers to have a unified cloud-native control and management experience with Azure services across the public cloud, on-premises infrastructure, and edge locations (Azure Arc brings customers a single pane of glass experience across premises.)
- **Launch of Dav4 and Eav4 AMD EPYC-based instances** – Making Azure the first major cloud provider to deliver AMD EPYC 7452 (Rome)-based Compute offerings
- **Azure Spot Virtual Machines (VMs)** – Increasing the range of cost-optimized consumption models available to Azure customers to consume Azure VMs and enabling up to 90% savings for customers that choose to use Spot VMs for their interruptible workloads
- **Azure Dedicated Hosts** – Dedicated physical servers to run Azure VMs for Windows and Linux in a single-tenant environment (Dedicated Hosts allow better support for multiple scenarios, including workloads with compliance, stringent physical resource allocation, or license optimization needs.)
- **Generation 2 VMs on Azure** – Enabling customers to deploy larger VMs with up to 12TB of memory, migrating generation 2 Hyper-V VMs more easily to Azure, and provisioning OS disks that exceed 2TB (This enables the new SAP HANA-optimized Mv2 VMs, supporting up to 416 vCPUs and 12TB of memory.)
- **Azure VMware Solution** – Providing customers a consistent experience across Azure public cloud and on premises for their VMware-based workloads
- **Proximity placement groups (PPGs)** – Allowing customers to specify colocation of virtual machines to minimize network hops and latency between VMs (This improves the performance of latency-sensitive applications, including distributed HPC workloads and multitier enterprise IT applications like SAP.)
- **Major additions to the HPC portfolio** – Including specialized compute offerings such as the Azure HBv2 Virtual Machine with the public cloud's only 200Gb HDR InfiniBand with AMD EPYC 7002-series CPUs and the Azure NDv2 Virtual Machine with the public cloud's only InfiniBand-equipped eight-way NVIDIA V100 Tensor Core GPU server for the most demanding HPC, artificial intelligence (AI), and machine learning (ML) workloads (The new AMD EPYC and Radeon-based NVv4 also give customers additional choice and flexibility, including offering partitionable partial GPUs, starting from one-eighth of a GPU.)
- **Expansion of Azure reservations to support 16 additional services, including Azure Premium SSD disk storage** – Giving better cost optimization options for the workloads on Azure
- **Introduction of new tools and services to support customer migrations to Azure, including the Azure Migration Program and the Azure Migration hub**

IN THIS WHITE PAPER

This white paper presents the Microsoft Azure compute portfolio and the broad range of computing capabilities available with Azure compute. It also discusses how Azure, combined with the broader Microsoft ecosystem, enables an easy path to infrastructure and application modernization for IT organizations.

SITUATION OVERVIEW

Enterprise IT adoption of public cloud infrastructure continues to grow, and workloads being deployed on public cloud now span a wide range of use cases – including business-critical operations, high-performance computing, and cost-sensitive short-term use cases. This movement of legacy as well as

new applications to the cloud has continued to push at the limits of capabilities on cloud platforms – with workloads demanding increasing levels of customization, flexibility, and control.

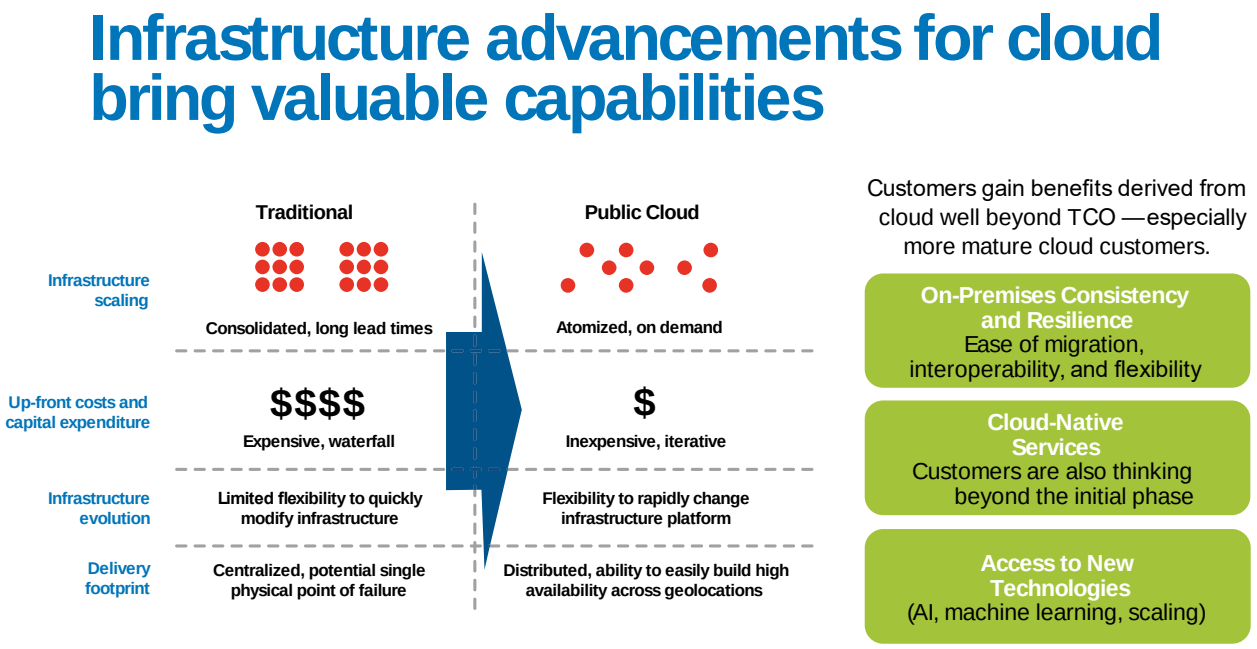
Momentum toward public cloud has been further accelerated by the 2020 COVID-19 disruption and the acknowledgement of the flexibility made possible through public cloud. Recent IDC surveys of enterprise IT decision makers indicate an increasing affinity to cloud computing for infrastructure needs, with a majority anticipating either no impact or an increase in cloud computing spend due to the COVID-19 disruption (seen *Impact of COVID-19 on IT Infrastructure Spending: Spotlight on Servers, External Storage, and Cloud Computing,* IDC #US46208920, April 2020).

The post-COVID-19 focus on public cloud has also been characterized by a shift in priorities and a renewed focus on resilience and cost. IDC's May 2019 *COVID-19 Impact on IT Spending Survey* reveals that the top IT priority among global IT decision makers has changed over the five months since the beginning of 2020. The top priority reported post-COVID-19 is "resilient business operations," a priority that was ranked fifth prior to COVID-19. Alongside, there is a renewed focus on cost, as businesses turn the focus from expansion to survival and from investment to optimization.

These shifts in priority, combined with the increasing use of public cloud, make it more important than ever that enterprises run workloads in a cloud computing environment that is optimized for the workload-specific needs. While business agility and access to new technology continue to be important priorities, the upcoming wave of cloud adoption from enterprises will be characterized by a renewed focus on cost control and optimization. The new IT infrastructure paradigm enabled by the public cloud is ideally suited to meet these needs (see Figure 1).

FIGURE 1

Cloud Changes the IT Infrastructure Paradigm



Source: IDC, 2020

In the post-COVID-19 era, cost optimization will again emerge as a top driver of cloud adoption. This includes initial adoption costs, impact on the total cost of operations, and cost of migration. To minimize the effort of migration, enterprises require consistency between their on-premises environment and the cloud, for DevOps as well as the runtime experience. Aside from cost impacts, enterprises continue to see the value of cloud-native and higher-layer services as they position themselves for digital capability–driven expansion. Running applications on public cloud IaaS makes it easy for the enterprise, whether central IT or lines of business (LOBs), to leverage technologies such as artificial intelligence, machine learning, and dynamic scaling.

Alongside cost, the strength of the ecosystem continues to be an important factor driving cloud provider selection, a growing driver of the move to cloud. As previously mentioned, cloud is seen as a source of access to new technologies, not just as a functional solution for infrastructure or IT. Adjacent enabling services such as machine learning and IoT are helping customers on their digital transformation (DX) journey, and such capabilities, and their easy availability, are no longer seen as differentiators but table stakes capabilities from the major global public cloud providers.

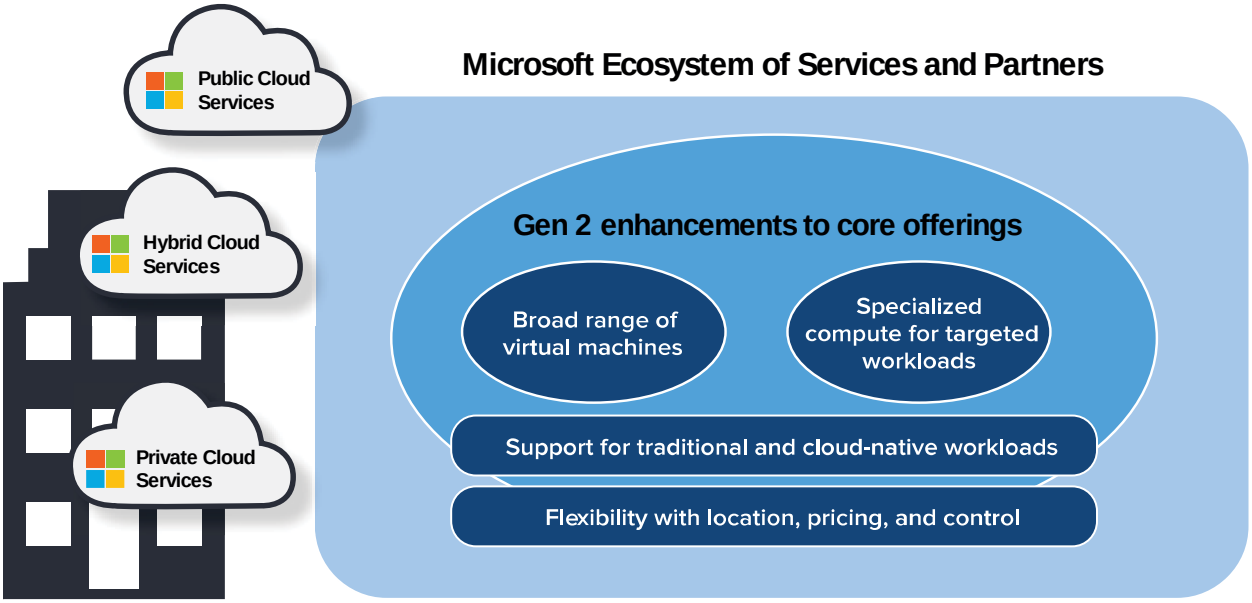
THE AZURE CLOUD COMPUTE PORTFOLIO

Microsoft released Azure, its public cloud application platform, in February 2010, but its public cloud IaaS focus really started ramping up in 2014. The Azure portfolio now includes an expanding range of public cloud, hybrid cloud, and edge deployment services, all delivering a consistent Azure experience.

At the core of the Azure cloud platform is a broad portfolio of compute VMs with support for both Linux and Windows OSs, with families for mainstream as well as targeted specialized compute needs and use cases (see Figure 2). This portfolio of cloud services is surrounded by a variety of additional capabilities such as containers and functions as well as advanced services for databases, AI and analytics, IoT, management tooling, and a marketplace of partner services. All these services are further supported by the global presence of Microsoft and its ecosystem of services, software, and partners.

FIGURE 2

Azure Compute Portfolio Within the Broader Microsoft Azure Ecosystem



Source: IDC, 2020

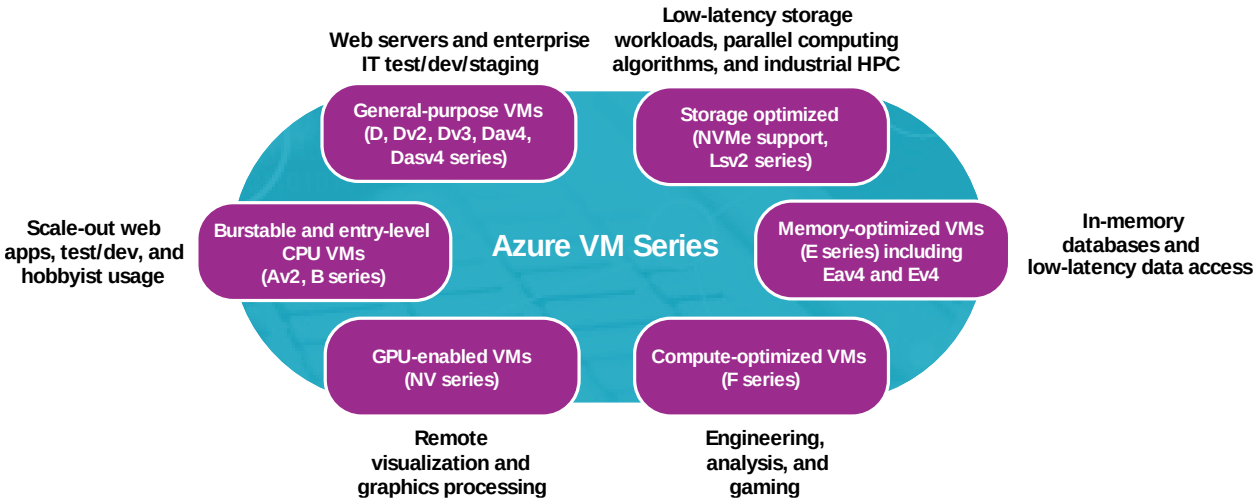
Compute for Entry-Level, General-Purpose, and Workload-Optimized Needs

Azure has maintained a steady pace of enhancement to the Azure compute portfolio, addressing initial gaps in the portfolio as well as bringing new enterprise IT-centric value propositions to market. Azure today offers a wide range of choices in terms of processors, configurations, and performance range so that enterprises can match the infrastructure to their specific workload needs (see Figure 3).

FIGURE 3

Microsoft Azure Compute: VMs for Every Workload

Broad array of VM families for mainstream compute needs – optimized options for common workloads and use cases



Source: IDC, 2020

Entry-Level VMs

Azure's basic and lowest-cost options, the A series and B series, consist of Linux and Windows VMs. These VM environments are applicable to low CPU utilization tasks such as code repositories and build servers for development and test, scale-out web applications, and low-traffic websites. B series allows customers to have a low baseline usage and periodically burst to higher CPU usage when needed. Windows and Linux versions of the B1S virtual machines are available as part of Microsoft's free try-before-you-buy account.

Mainstream VMs

These are VMs targeted to deliver optimized processor core, memory, and storage performance configurations for a wide range of typical enterprise IT use cases.

Intel VMs for general-purpose, memory-optimized, and compute-optimized workloads are delivered through the D series, E series, and F series of VMs. The baseline D series offers a general memory-to-virtual CPU configuration for general-purpose test, development, and staging related to enterprise-level workloads. The E series has higher memory-to-CPU ratios, uses hyperthreaded Intel Cascade Lake and AMD EPYC processors, and is optimized for memory-intensive workloads including relational database servers such as SAP or Microsoft SQL Server that require low-latency data access or in-memory analytics. The F series has higher CPU-to-memory ratio and uses the high-performance Skylake or Cascade Lake processors, optimized for compute-intensive workloads such as gaming, engineering, analytics, and high-demand web servers.

Azure also delivers a growing range of AMD EPYC VMs, building on Microsoft's early engagement with AMD. Azure has continued to expand on its AMD-based virtual machine offerings. The AMD EPYC 7551-based Lsv2 series is designed for storage-intensive use cases and uses NVMe-connected

SSD local storage to deliver high IOPS and throughput for the workload. Azure also delivers specialized AMD EPYC-based virtual machines like the HBv2 and NVv4 targeting high-performance use cases.

In late 2019, Azure also became the first public cloud provider to deliver AMD EPYC 7452-based VMs on the cloud, with the Dav4 and Eav4 series. The Dav4 series is designed to be a cost-competitive alternative for general-purpose workloads, supporting 2-96 vCPU, up to 384GiB of RAM, and usable alongside Azure's high-performance SSD disks. The Eav4 series is designed to meet memory-optimized workload needs and supports up to 672GiB of RAM.

With these, Azure continues to extend the range of cost-competitive compute options available to enterprise IT organizations for mainstream workload requirements.

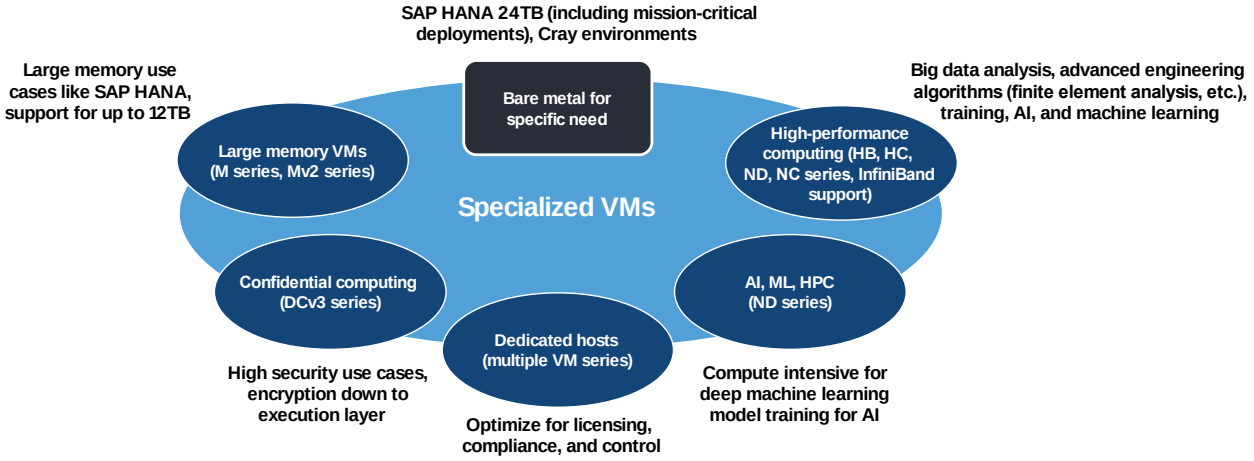
Specialized Compute VMs and Infrastructure

The Azure compute portfolio also offers a multitude of VM options designed for specific high-performance computing and high-performance data analysis (HPDA) needs (see Figure 4).

FIGURE 4

Microsoft Azure Compute: Specialized VMs and Infrastructure

Targeted portfolio of specialized compute offerings for specific needs and use cases



Source: IDC, 2020

The N-Family and H-Family for AI/ML and High-Performance Use Cases

Azure N-family offers customers varying configurations of virtual machines enhanced with NVIDIA GPU coprocessors. These include the NC, ND, and NV series. The NC series, built with a range of NVIDIA GPUs (K80, P100, and V100), is designed for compute-intensive workloads or HPC multithreaded parallel processing applications using GPU acceleration. The ND series focuses on training and inference scenarios for deep learning. The NDv2, released at SC 19, supports 8 NVIDIA V100 GPUs, with 32GB of high-bandwidth memory (HBM2) per GPU, and 100GB EDR InfiniBand interconnectivity, alongside a Xeon Platinum 8168 processor with up to 40 non-hyperthreaded cores and 672GiB of system memory. NDv2 has been demonstrated to support deployments of up to 800

NVIDIA V100 GPUs interconnected on a single InfiniBand network and is designed to meet the most demanding HPC and AI/ML workloads.

The NV series is built to support powerful remote visualization workloads and other graphics-intensive applications. The NV and NVv3 series are built on NVIDIA Tesla M60 GPUs and Intel Xeon CPUs. The latest enhancement to this includes the NVv4, built on AMD Radeon MI25 GPUs and AMD EPYC 7742 processors, supporting up to 16GB of HBM2 memory. The NVv4 also allows customers to consume fractional GPU sizes starting from one-eighth of a GPU, increasing the choice and cost optimization opportunities for customers.

The H series is tuned for tightly coupled parallel computing, connecting HPC clusters via the Message Passing Interface (MPI). These VM environments are tuned for HPC and HPDA workloads such as weather forecasting, statistical and financial modeling, advanced engineering algorithms such as finite element analysis and seismic processing, Monte Carlo simulations, and genomics, as well as AI and machine learning. The HBv2 series is designed to deliver supercomputer-class performance in a cost-effective manner. This makes it a compelling choice for real-world HPC workloads. The HBv2 comes with 120 non-hyperthreaded AMD EPYC 7742 cores, 4GB of system memory per core, 350GBps memory bandwidth, and 200GB HDR InfiniBand interconnectivity.

The HBv2 delivers 4 teraflops of double-precision operations, which, combined with the large memory bandwidth, makes it well positioned to serve typical real-world HPC use cases like fluid dynamics and finite element analysis. Sample computational fluid dynamics (CFD) simulations, published at the Microsoft web portal, have highlighted HBv2's ability to scale efficiently to a high degree of parallelism. In addition to these, HBv2 delivers the same pay as you use and high scalability of other Azure compute offerings. Partly owing to these, HBv2 has been referred to by reports as the "most affordable yet powerful HPC server from Azure."

Bare Metal and Large Memory VMs for Mission-Critical Workloads

Azure offers access to single-tenant bare metal servers for workloads that require specialized infrastructure — such as a dedicated Cray supercomputing platform. This gives enterprises access to specific supercomputing resources as needed in an Azure environment. Azure also offers customers purpose-built infrastructure for SAP HANA, supporting up to 24TB per node for scale-up SAP HANA environments. These are complemented by the SAP-qualified M-series virtual machines, designed to support large in-memory data management systems, with support for up to 12TB of memory per VM.

Confidential Computing for the Most Security-Sensitive Workloads

The expanding use of public cloud has also driven a steady increase in demand for higher security capabilities to enable customers to confidently deploy security- and compliance-sensitive use cases on public cloud. While data-at-rest encryption and data-in-motion encryption protect the data when it is outside the computing environment, these do not cover data that is in use in the processor. Azure confidential computing allows customers to encrypt data in use within the processing environment through the use of isolated trusted execution environments (TEEs). This protects the data from threat sources like malicious inside users, exploits that may take advantage of vulnerabilities in the hypervisor or OS, or even third-party partners that may not have privilege to access customer data.

Azure confidential computing uses hardware encryption enabled by Intel SGX technology for the creation of TEEs. Microsoft is a founding member of the Confidential Computing Consortium (CCC), formed to create an industry standard and accelerate availability and adoption of confidential computing. Microsoft will also be contributing to the Open Enclave SDK, being developed to create a unified development approach across multiple underlying hardware-based technologies that can

support creation of TEEs. Azure confidential computing is available to customers through the DC-series virtual machines.

Azure VMware Solution

As enterprises increase their use of public cloud alongside traditional infrastructure, there is a growing realization of the value of consistency across these environments — especially during the initial adoption or workload transition period. Azure's early integrations with VMware were intended to address this need. In 2020, Azure introduced its next generation of VMware enablement with Azure VMware Solution. Azure VMware Solution is a natively delivered Azure offering, verified by VMware, that allows customers to operate workloads on Azure using the same VMware tools and workflows used with their on-premises VMware environment.

Azure VMware Solution allows customers to provision single-tenant bare metal hosts within the Azure infrastructure, with support for vSphere, vSan, and NSX, without any of the overhead of procuring or operating the servers. They can then launch virtual machines in these hosts, just like they would with any other host server, using the VMware tools or API. Customers can also use the broader Azure cloud services, including databases, analytics, and AI/ML services, alongside their VMware VMs. At the time this document was written, Azure VMware Solution is available in preview for customers.

Azure for Cloud-Native and Traditional Applications

Enterprises are now using cloud for both traditional and cloud-native use cases. Commonly seen adoption paths include new cloud-native projects built first in the cloud, modernization of existing applications, and migration of traditional Windows and Linux workloads using a "lift and shift" approach.

Azure for Cloud-Native Compute Needs

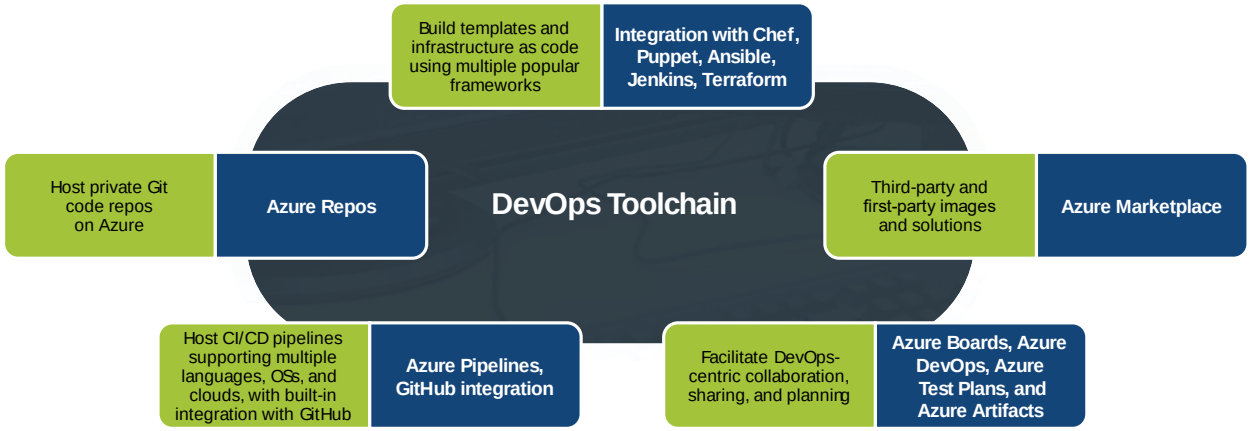
Azure has a growing portfolio of cloud-native tools and methodologies supporting next-generation cloud-native application development. Azure Kubernetes Service (AKS), Azure Container Instance, and third-party container orchestration are available for managed and self-operated container clusters. Azure has public and private container registries for sharing and reuse of container images, with Azure's Container registry and Docker Hub support. Azure Functions and Azure Logic Apps support the development of serverless and event-driven architecture-based applications. In addition, higher-layer abstractions like Azure Service Fabric and Azure App Service are available to support microservices-based applications and hardware-abstracted native platform-as-a-service (PaaS) development. Microsoft is a Platinum member of the Cloud Native Computing Foundation and continues to invest in Azure cloud-native capabilities. These are further enhanced by the rich Azure DevOps toolset.

Azure DevOps delivers Azure-native tools and external extensions that offer developers a choice of DevOps-based pipelines and collaboration tools (see Figure 5). Users can build templates and infrastructure as code using multiple popular frameworks (Chef, Puppet, Ansible, Jenkins, and Terraform) and create repositories and host private Git code with Azure Repos. With Azure Pipelines and GitHub integration, developers can host continuous integration and continuous deployment (CI/CD) pipelines supporting multiple languages, OSs, and clouds, with built-in integration with GitHub. Tools such as Azure Boards, Azure Test Plans, and Azure Artifacts facilitate centric collaboration, sharing, and planning for DevOps. In the Azure Marketplace, numerous other third-party and first-party images and solutions are available to help developers utilize the Azure environment.

FIGURE 5

Microsoft Azure Compute: DevOps Toolchain

Azure-native tools and external extensions – enabling choice of DevOps-based pipelines and collaboration tools



Source: IDC, 2020

Azure for Traditional Windows and Linux/Open Source Workloads

IDC data shows that Windows servers account for nearly half of the server OS install base worldwide. Microsoft has actively increased its focus on Linux in the past five years and is now a significant contributor to the Linux kernel, according to the Linux Foundation. Microsoft has been a member of the Linux Foundation since 2016.

Azure VMs can be deployed with either Linux or Windows. According to Microsoft, nearly half of Azure deployments run various flavors of Linux – including RHEL, Ubuntu, Debian, SLES, CentOS, CoreOS, and Oracle Linux. Azure offers turnkey Linux image solutions from the Azure Marketplace. The GitHub acquisition increases Microsoft's integration with the open source community, and Microsoft is a member of several open source foundations (Node.js Foundation, Cloud Native Computing Foundation, and Cloud Foundry).

On the Windows side, Microsoft has worked hard to encourage its massive installed base of Windows Server customers to make use of the Azure cloud in a variety of ways and is providing attractive on-ramps to encourage adoption. For example, Microsoft offers Azure Hybrid Benefit to Windows Server and SQL Server customers, giving them the ability to use its Windows and/or SQL Server licenses on Azure at the lower Linux IaaS rates. There is no cloud environment that is better for the Windows Server OS than Azure, which has a deep engineering understanding of optimizations needed for Windows and Microsoft enterprise applications.

Easy Migration to Azure

IDC's 2019 *IaaSView Survey,* a survey of enterprises using public cloud IaaS, shows that 86% of respondents report cloud migration as one of their top 2 priorities around cloud over the next year. Azure offers assistance through all major phases in the migration journey – assessment of the application portfolio, migration, optimization, and management of migrated workloads. One way in which this is enabled is through targeted tools – such as Azure Migrate for VM and physical server migration, Azure Site Recovery to execute a failover migration, Azure Database Migration Service, and

Azure Data Box for data movement. These are complemented by additional assistance through the Azure migration center, Azure Migration Program, and FastTrack for Azure. Azure also works with its broad partner network to deliver these in an optimal manner to end customers worldwide.

Access to easy and cost-effective migration tools has been repeatedly highlighted in customer surveys as one of the top factors influencing customers' choice of public cloud partners. Availability of these tools will reduce customers' challenges as they increase their usage of Azure.

Generation 2 Virtual Machines on Azure

Microsoft generation 2 Hyper-V virtual machines were initially introduced in 2016 with Windows Server 2012 R2. Support for generation 2 VMs on Azure was announced in late 2019. While all of the generation 2 features on Hyper-V are not yet available on Azure, generation 2 VMs on Azure do offer an easier and more direct migration path for generation 2 VMs on premises, as well as critical functional capabilities for advanced use cases. These include support for Unified Extensible Firmware Interface (UEFI)-based booting; Intel SGX, which underpins the confidential computing capability; and SCSI disk controllers. Generation 2 VMs allow for larger VM sizes with up to 12TB of memory and OS disk volumes larger than 2TB.

A current listing of virtual machine series supporting generation 2 VMs can be found on the Microsoft Azure web portal.

Enhanced Placement Control, Cost Optimization, and Automation

Azure Dedicated Host service delivers a dedicated physical server for the customer's virtual machines. Customers can host a mix of Windows or Linux virtual machines in the Dedicated Host, with the assurance that resources on that physical server will not be shared with any other customers. The underlying physical server for the dedicated host is just another server in the Azure infrastructure fleet, connected to all Azure services and resources and manageable through Azure APIs. Customers can launch Azure VMs on the provisioned Dedicated Host using the same workflows and APIs they would use to launch any other Azure VMs. Dedicated Hosts also support the maintenance control feature, using which customers can control all host updates, including rebootless updates, within a 35-day window, allowing much higher predictability around maintenance events for the VMs.

Dedicated Hosts provide customers a higher level of control over the physical resources, allowing for better planning for resource allocation, maintenance planning, and VM placement. Dedicated Hosts also allow better alignment with certain compliance requirements, where there may be restrictions or requirements around the level of isolation required for data or applications. Besides the control and compliance benefits, Dedicated Hosts can also drive cost of ownership improvements for customer workloads, through optimized use of existing software licenses. This may be through optimization of Windows licensing using programs like Azure Hybrid Benefit, or for non-Windows licenses through the use of bring-your-own-License (BYOL) programs.

Proximity placement groups allow customers to colocate their desired Azure compute resources to minimize network latency between the resources. PPG is a new deployment constraint that customers can specify when creating new VMs, availability sets, or VM scale sets. VMs launched with the same PPG attributes are provisioned in the same datacenter, allowing lower network hops and through traffic latency between these VMs. For multi-VM applications, such as typical three-tier enterprise applications, low latency between the tiers is a critical determinant of performance and user experience. Proximity placement groups allow customers to create resources with predictably low latencies between resources. This enables deterministic performance for distributed and multitier

workloads, including SAP workloads, high-performance computing clusters, and tightly coupled scale-out applications.

Cost Optimization Through Broad VM Consumption Models and Ephemeral OS Disks

Cost optimization continues to be an important focus for IT organizations, and the current COVID-19 disruption has increased this focus on cost. Azure allows multiple levers for customers to manage their Azure usage costs. The most direct among these is through the multiple consumption model options available — with on-demand, reserved, and spot pricing. Based on the customers' compute needs, they can choose to commit to one or three-year reservations, driving savings of up to 72% for Linux VMs and even higher when considering Azure Hybrid Benefit. Azure reservations also now extend beyond virtual machines, supporting 16 additional Azure services including Premium SSD disks, Databricks, and SQL Data Warehouse — offering the ability to use the reservation-based consumption model to control broader costs beyond control.

In December 2019, Azure also announced the introduction of Spot VMs, which is unused Azure capacity offered to customers at a "maximum price" set by the customer. If the maximum price specified is higher than the Azure-set current spot price for the VM, the VM is launched. Once the current spot price exceeds the maximum price, the VM is deallocated and the workload is terminated with a 30-second notice. Spot VMs are ideal for running interruptible use cases like batch processing and other stateless processing applications, as well as certain compute-intensive workloads that require large capacities for short bursts of time.

Another way in which Azure allows customers to make workload-specific optimizations is through Ephemeral OS disks. Ephemeral OS disks are created on the local VM storage and not persisted to the Azure Storage back end. This makes Ephemeral OS disks ideally suited for workloads that are tolerant to potential loss of data due to individual VM failures and do not require data persistence. The local nature of the Ephemeral OS disks also allows them to support faster boot and provisioning compared with instances using Azure Managed Disks. Ephemeral OS disks work with Marketplace and custom images as well as Shared Image Gallery, making them well suited for scalable use cases with high error tolerance for individual VM failures. Ephemeral OS disks are offered to customers at no cost, allowing customers to make trade-offs across cost, provisioning lead time, and data persistence requirements. Spot VMs and Ephemeral OS disks are now generally available on Azure.

Azure Virtual Machine Scale Sets for Automated Application Resilience

Azure's Virtual Machine Scale Sets (VMSS) provides customers an automatically scaled group of VMs, with configurable parameters to trigger an increase or decrease in the number of virtual machines. VMSS allows customers to easily build in agility and resilience into scale-out applications that are distributed across multiple VMs. Recent enhancements introduced let customers create Virtual Machine Scale Sets using VMs from different VM series, allowing richer configurations within the scale set. Customers can also deploy scale sets for high availability by deploying across multiple fault domains, ensuring higher resilience and reduced exposure to single point of failures.

VM Scale Sets can now be configured for automated provisioning using images from Shared Image Gallery, allowing for faster scaling and faster distribution of updated custom images for VM Scale Sets. Other recent enhancements include support for notifications preceding termination and the ability to designate VMs from being terminated when the scale set is being decreased in size. These allow customers to have higher agility and control with their VM Scale Set deployed workloads.

AZURE COMPUTE BENEFITS

In the past three years, Azure has made significant enhancements to its cloud IaaS portfolio, addressing gaps as well as bringing new enterprise IT-centric value propositions to market. Key to these enhancements are the underlying tenets of the portfolio — compute for all workloads, cost-effectiveness, security, and a true hybrid cloud environment and management services for a diverse cloud environment.

Compute for All Workloads

Azure is a core focus of Microsoft's cloud portfolio, with the vision to provide an Azure solution regardless of the service need. We have already described the breadth of Azure's compute portfolio, which ranges from sandboxes to supercomputing. Microsoft offers Azure-based solutions for a myriad of workload situations. Given the company's pace of innovation, it also means that customers frequently have access to advanced features.

Cost, Security, and Flexibility for Enterprise Compute Needs

Cost

The cloud environment reduces risk by minimizing up-front investment and providing flexibility through the opex consumption model. Additional cost savings can come from fine-tuning and optimizing the environment to match workload needs. This is enabled by Azure's broad range of targeted compute offerings. These are further complemented by the flexibility in consumption models, across on-demand, reserved, and spot virtual machines.

In addition to optimized resources and consumption options, Azure allows customers to take advantage of Microsoft license mobility programs like Azure Hybrid Benefit. With Azure Hybrid Benefit, users bring their own Windows Server or SQL Server on-premises software license to the cloud, eliminating the need for multiple software licenses for each of the VMs used. According to Microsoft, customers that opt for RIs with a three-year term can experience savings of up to 80% compared with the pay-as-you-go approach. These are further enhanced for Microsoft Software Assurance (SA) customers by the extension of Azure Hybrid Benefit to new offerings like the Azure Dedicated Host and the 180-day dual-use license provided to customers migrating Windows Server or SQL Server licenses to Azure to cover the transition period of running simultaneously on both platforms.

Last but not least is the rich portfolio of management and cost oversight tools offered within Azure. These tools provide the ability to plan and budget usage, using tools like the TCO calculator; guard against low CPU utilization while maintaining application performance, using guidance from Azure Advisor; and continuously monitor and optimize cloud usage costs, using Azure Cost Management. Cumulatively, these offer customers an automated and built-in framework to run their workloads on Azure in a continuously optimized manner.

Security

Security is often cited as an inhibitor to cloud adoption, yet cloud providers such as Microsoft are uniquely capable of embedding security into their systems.

Rigorous security and privacy standards are at the core of the Azure offering, with a focus on privacy, transparency, and compliance. Azure has built-in security controls and intelligence and is compliant with over 90 industry certifications such as the General Data Protection Regulation (GDPR), ISO 27001, HIPAA, FedRAMP, SOC 1, and SOC 2. Users have the ability to exercise governance via software tools such as Compliant Manager, which provides a GDPR readiness score; Azure Policy;

Azure Security Center, a security management and threat protection solution dedicated to hybrid cloud workloads; Azure Monitor, which uses advanced analytics and machine learning to monitor applications, infrastructure, and network; and Azure Backup for easy and integrated protection of data on the Azure platform. Azure Site Recovery offers built-in disaster recovery as a service (DRaaS). Azure-native management tools manage applications throughout the asset life cycle. AI-enabled Azure Log Analytics provides continuous monitoring, status, and security. Microsoft's global and redundant datacenter deployments reduce the risk of service interruptions. The company currently has 60+ global datacenters across regions, offering data residency choices in a broad set of locations. Microsoft Azure is the only major public cloud platform with built-in VM disaster recovery capability through replication to a secondary region. Microsoft's strong security posture comes from a deep company history with enterprise IT buyers and users.

Flexibility: Seamless Hybrid and Multicloud Environment

IDC's 2019 *IaaSView Survey* found that 52% of companies already have a hybrid cloud infrastructure and that 32% expect to implement a hybrid environment within the next year. Cloud service providers are also adapting to the fact that theirs is not the sole environment for many customers. Microsoft is the first public cloud vendor with a comprehensive offering across public cloud and on-premises environments while meeting compliance and performance requirements with one consistent framework (Azure). Microsoft recognizes that more organizations are pushing for hybrid and multicloud solutions than ever before and is clear about facilitating a Microsoft-only solution both on premises and off premises as well as providing solutions — such as Azure Arc — for multiplatform environments.

Microsoft Azure Stack provides an Azure-consistent cloud-native environment on premises and at the edge. With Azure Data Box Edge, customers can quickly process data onsite before transferring the data to the Azure cloud for further analytics and distribution. Depending on the onsite scenario, Microsoft provides Azure Stack implementations of different sizes. Azure Stack includes consistent tools and integrated management capabilities such as Azure Resource Manager, Azure Services, Azure DevOps, containers and serverless functions, Linux, and Windows. Qualified hardware partners include Dell, EMC, HPE, Lenovo, Huawei, Terra, and Avanade. Microsoft provides Service Bus for asynchronous message brokering between applications and services in a hybrid cloud environment.

The consistent and smooth linkage via Azure Stack of customer-premises systems with the off-premises Azure environment enables customers to address increasingly important multicloud requests in a single cloud provisioning and orchestration model. IDC believes that a multipremises environment that encompasses local (on-premises) and public cloud resources within a single cloud platform is an important enabler for sustained digital transformation, and the Azure Stack is an important differentiator in this direction.

Azure Management Capabilities

Key to any controlled IT environment are configuration, operations, and workload management capabilities that include monitoring and optimizing as well as ensuring a secure infrastructure. Microsoft provides a plethora of enterprise-level tools that allow administrators to manage its Azure cloud resources. These management tools provide for configuration (Azure Resource Manager, Advisor, Portal, Traffic Manager), migration (Azure Migrate, Database Migration Service), monitoring (Azure Monitor, Log Analytics, Application Insights), governance (Azure Policy, Cost Management), and security (Azure Backup, Site Recovery, Security Center).

For example, Azure Resource Manager templates present infrastructure as code and provide templating and reuse across defined environments such as development, test, and production.

Modernization and the Azure Ecosystem

Public cloud is increasingly seen as a source of access to new technologies and higher-layer services for modernization of applications, not just as a functional solution for infrastructure or IT. A vibrant ecosystem is a growing component of cloud adoption and usage.

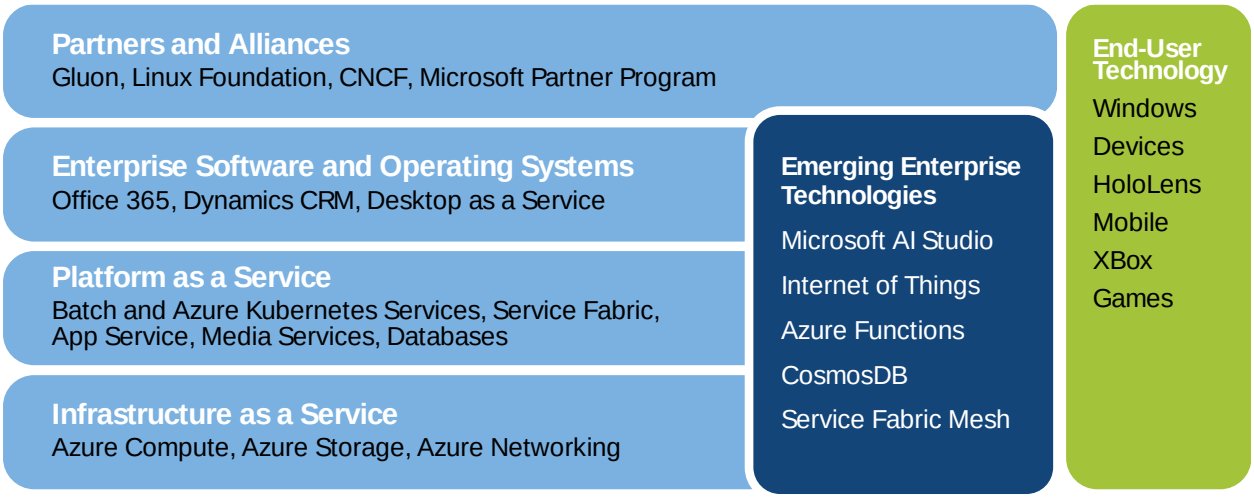
Azure provides a methodology, native and partner tools, and an ecosystem of business partners to help customers modernize their existing on-premises applications and migrate them to the cloud. Azure Migrate offers a structured process to assess, move, optimize, and secure workloads in Azure, with tools to analyze applications and discover dependencies. Microsoft also offers VMware virtualization options on Azure to migrate and modernize applications.

The Azure Marketplace contains over 4,000 applications that support a variety of use cases – third-party engines, IoT, backup, and recovery. Azure partners include ISVs that contribute applications that run in the Azure environment (such as NetApp's enterprise-grade network file system [NFS] service as well as hundreds of other applications) and systems integrators (SIs) positioned to help customers take full advantage of the Azure cloud and all its capabilities.

The strength of the Microsoft cloud ecosystem is another key source of benefit and productivity for customers as they progress on their cloud adoption journey (see Figure 6).

FIGURE 6

Microsoft's Cloud Ecosystem and Presence Across Technology Markets



Source: IDC, 2020

Azure AI and the Broader Microsoft Intelligent Cloud Ecosystem

Inherent advantages for Microsoft are its dominant presence in workplaces around the globe and its ability to offer a true hybrid cloud solution with Azure Stack on-premises connecting to Azure public cloud services. In addition to these, Microsoft recognizes the urgency of having an artificial intelligence strategy for its developer community and has introduced a broad range of AI functionalities for easy consumption by developers.

The Azure AI platform contains all the components for incorporating AI: data prep, build and training of models, and deployment. Azure offers a broad portfolio of services – prebuilt, custom, conversational,

and data science VMs for data analytics, machine learning, and AI. Prebuilt APIs include trained models for images/videos, speech and language, and search. Azure also offers a bot-building environment where developers can build and deploy intelligent conversational applications or chatbots. Through these, Microsoft ticks all the boxes for AI tools, frameworks, and scalable infrastructure.

Customer Success Stories

Deep Dive: On Premises to Azure Public Cloud in Six Months

A government agency in the United Kingdom launched an IT modernization exercise with Microsoft Azure public cloud as the public cloud partner to facilitate this transformation. The existing IT deployments and operations at the agency were completely on premises, and the team had limited assets or experience with public cloud at the onset. A large part of the first phase of this effort was identifying the right compute service to match application needs and moving each application to its optimal compute environment through a "lift and shift" type of exercise.

The team was able to complete the planning and execution of the entire migration within six months, the latter part (execution of the actual migration) taking only weeks to complete. Key factors highlighted by the customer as enablers for the smooth migration experience include:

- **Broad set of compute options with Microsoft Azure.** The broad range of compute options allowed the team to easily map application requirements to target compute services, without making compromises in application needs or unnecessary overprovisioning.
- **Availability of migration services within the Microsoft ecosystem and through partners.** The customer used both Microsoft Azure Site Recovery and Microsoft partner Velostrata to aid in the migration. The automated migration capability was a critical factor in the rapid, successful migration of applications.
- **Existing experience with Microsoft tools.** The team had limited experience with Microsoft Azure, but it was familiar with Microsoft development and asset tracking tools. This experience allowed a portion of processes to be transitioned into the public cloud environment with few or no changes.
- **Expertise available through Microsoft partners in the planning process.** Besides the technical capabilities delivered by Microsoft and its partners, the customer was able to obtain strong guidance for planning its migration into Azure from its existing professional services partner. This complemented the feature sets available and allowed the team to quickly address internal skill set gaps as needed during the planning and execution of the migration.

The first phase of the transformation has been a strong success at the agency. It has also provided confidence and impetus to set the goal higher for the next phases of the digital transformation journey.

Other Recent Customer Successes: Speed and Scalability with Azure Public Cloud

A major tax services company in the United States selected Azure as its partner to unify its disconnected data sources and silos, with the intent of delivering a unified multichannel experience to customers. Through the use of SQL Server 2017, Azure SQL Database managed instances, and broader Azure data services, the customer has been able to redesign its back-end data store to improve performance and scalability while creating a unified platform for innovation.

Another recent success story comes from a major insurance company in the United States that wanted to modernize its sales data infrastructure, which was originally on a departmental mainframe. The intent was to drive agility and increase competitiveness. The organization took the path of rewriting the

application as a .NET application and looked to maximize leverage of services on the Azure hosting, data services, and monitoring needs of the application. This transformation has allowed it to decrease time to market for new products, hire new talent more easily, and have access to new technologies such as AI for new-initiative experimentations.

Both illustrate how Azure compute and the broader capabilities in the Azure ecosystem empowered customers to achieve modernization goals and to move to a more agile and responsive IT platform through the use of Azure.

Assisting Customers Through COVID-19 Disruption

The recent COVID-19 disruption highlighted the importance of access to a resilient and scalable IT platform, particularly for high-performance computing needs to support simulations and analysis to better understand the situation. Microsoft has been on the forefront of supporting these needs. This includes ensuring availability of the Azure HPC offerings for such needs, publishing a clear stance on prioritizing COVID-19-related workloads during this time, and participation in industry consortiums to lead the technology dimension in this effort.

Key among these industry consortiums is the COVID-19 High Performance Computing Consortium, which is a coordinated partnership across major technology providers and the government to provide HPC expertise to address the COVID-19 response using the XSEDE framework to share computing resources. Microsoft has been an active member in terms of offering the HPC resources, as well as participation by Microsoft researchers and data science experts to assist with the effort.

In addition, Microsoft has brought to customers a clear message on areas it can assist — the ability to work from anywhere, support for healthcare research, and the ability to accelerate insights with Azure HPC and Azure AI. Each of these are supported by a strong portfolio of specific tools and offerings and by Microsoft expertise to accelerate value from these capabilities. This clarity of focus is helpful for customers as they navigate the uncertainty introduced by this disruption and look for partners to optimize their operations.

CONCLUSION

Organizations are increasingly looking to cloud to enhance their IT capabilities and to gain access to the resilience, new technologies, and cost flexibility offered by cloud platforms. Critical to achieving these goals is selecting a cloud provider that can address the specific needs of the organization's use cases and workloads and provide an easy path to the cloud for the organization and its workloads. Microsoft Azure builds on Microsoft's decades-long experience with enterprise IT organizations and is designed to deliver on both these dimensions.

Azure has maintained its pace of portfolio expansion over the past two years and introduced new capabilities to support performance improvements, cost optimization, and control. Azure compute and the broader Azure portfolio continue to expand, allowing customers to easily bring to Azure an ever-growing portfolio of workloads and use cases in an optimized manner. Alongside, Microsoft has invested in its strong partner network to meet the specific needs and challenges of its broad customer base and to make consumption of Microsoft Azure easier for customers. With these, Microsoft Azure delivers to the market a compelling portfolio of cloud computing services to "meet customers where they are" on their DX journey.

Engage your Microsoft partners today and explore how Microsoft Azure can accelerate your organization's digital transformation journey while minimizing additional investments.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com