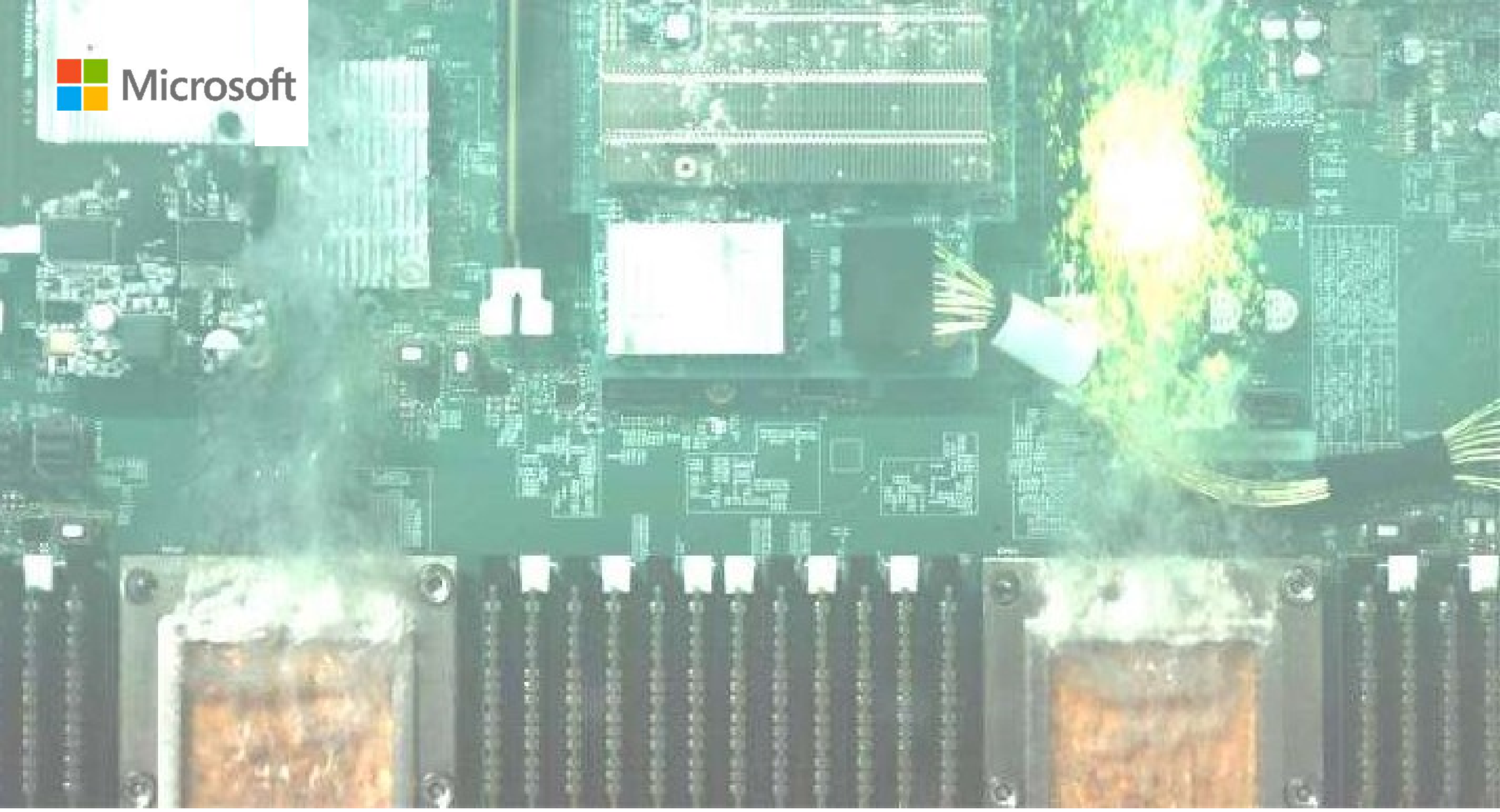

Azure for the semiconductor industry

Silicon-design workloads with electronic design automation software

Azure Silicon Industry Team

agisilicon@microsoft.com

June 2020

Contents

List of figures

Authored by the Azure Silicon Industry Team: Prashant Varshney, Mujtaba Hamid, Bella Chan, Preeth Chengappa, and Sonia Blouin with contributions from Steve Roach, Andy Chan, Michael Requa, Bernie Behn, and Giancarlo DiPasquale. Edited by Nanette Ray.

Executive summary

The connected world is a big business, and semiconductors are the underpinnings of every industry segment—from communications, computing, and automotive to the Internet of Things (IoT) and media and entertainment. The insatiable demand for semiconductors has pushed global revenues past the $450 billion mark. Semiconductor industry analysts forecast a 12 percent or higher year-on-year growth rate that, in short order, will push sales to $500 billion and beyond.[1] Relentless competition for those revenue dollars drives schedules, productivity, and innovation.

Shorter product refresh cycles, frequent evolution of standards, and the constant need for more performance are compressing the design cycle and time to market. Designs are growing more complex, and organizations need better design flows and comprehensive validation at every stage of development. The new process geometries also require massive compute power to address the process variability at such tiny geometries.

These trends—demand, speed, complexity, and agility—call for a reassessment of how silicon design is done from a toolchain and infrastructure perspective.

Microsoft is working to improve the complex electronic design automation (EDA) software landscape, boost productivity, optimize resources, and speed up time to market. We work closely with foundry partners and EDA vendors to develop finely tuned solutions that run on Azure high-performance computing (HPC) infrastructures. Our solutions include targeted reference architectures that optimize each EDA workload in the cloud. We've seen firsthand how semiconductor companies can take advantage of Azure solutions to achieve better quality designs, innovative solutions, and faster product delivery.

Azure also frees developers from the limitations of infrastructure performance and availability. Development teams can focus on running the right number of iterations, simulations, and regression tests in smaller windows to deliver greater functionality, higher quality, and more customizations. In addition, Azure supports teams across the development life cycle with agile DevOps tools and reusable best practices that ease the transition to the cloud platform.

Compared to the on-premises experience, Azure also offers ways to optimize the cost of ownership of cloud resources while maintaining or enhancing the performance. Deep understanding of the silicon industry helps us develop and deliver unique financial and ownership models that make Azure not just a viable—but a cost-optimal—solution. In addition, with the help of our long-standing partnership with the EDA vendors, Azure teams help drive the optimization of resource use directly into the EDA tools, as well as support innovation in EDA licensing models.

This document shows why Azure is the right solution for silicon design. It focuses on the tools and technologies that are already being used successfully by some of our biggest customers in the semiconductor industry and that we have validated with our industry partners.

[1] "Semiconductor market size worldwide from 1987 to 2020." Thomas Alsop. March 30, 2020.

Part 1: Silicon design in Azure

Silicon design is a highly complex process. Dozens of EDA tools are used in a typical chip design workflow. Each step in the workflow (workload) has unique infrastructure requirements. Each workload and its corresponding EDA tool poses its own challenges, from the intricate software algorithms governed heavily by Amdahl's law to the technologies used to manage licenses, schedule jobs, and distribute files across networks. To run efficiently, companies spend a lot of time performance-tuning their on-premises infrastructure.

EDA also constitutes big compute workloads. From a cloud-computing perspective, the security, reliability, and scalability that Azure offers as an HPC platform is a natural fit for EDA. Plus, Azure is always evolving. Organizations can choose from a growing set of cloud services that free developers to build, manage, and deploy applications on a massive, global network using the tools and frameworks they prefer.

For semiconductor companies considering a move to the cloud, Azure offers:

- A highly performant infrastructure capable of supporting production-level silicon design work.
- Strategies for migrating EDA toolsets to an Azure infrastructure optimized for performance and cost.
- A rich partner ecosystem, including semiconductor houses, foundries, tool vendors, and system integrators.
- Solutions for solving common industry impediments, such as product gaps, security, and cost of ownership.

Silicon design workflow on Azure

Each step in the complex silicon design process—from the initial specification of the front-end design to the final GDSII at tapeout for the back-end design—has a dedicated set of EDA tools. Azure services and tools can help optimize the design and collaboration environments used in these workflows, so teams get the most efficient results. Azure also offers mechanisms to help organizations choose the right combination of performance and cost for these workloads.

Front-end (logical) design

The front-end design phase flows from a specification to a logically validated design using software simulation. Like software research and development, this phase is code-heavy (primarily Verilog or VHSIC Hardware Description Language), with some block-level simulation and debug cycles. This workflow benefits from Azure DevOps tools and services, which improve the design cycle by fostering team collaboration and agile practices.

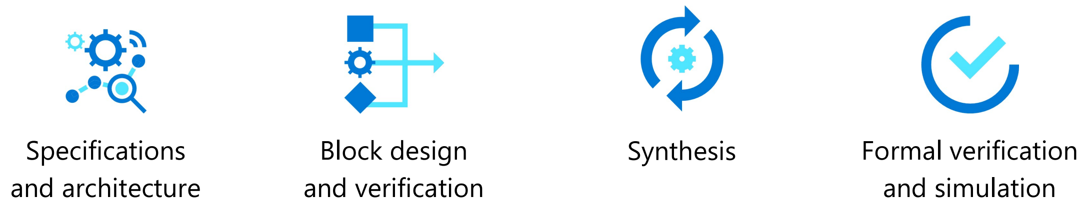

Figure 1. Steps in the front-end design phase use software simulation.

Following block design, the next step in the logical phase is to verify the design by running it on a large compute grid. With the support of a job scheduler, a design is distributed to available compute resources across the entire grid—with up to thousands of cores—to overcome the performance limitations of a single computer (node) running large, complex operations. Azure, with its near-infinite compute resources, can accelerate the verification process by shortening each job's runtime and by allowing more jobs to run in parallel.

Design synthesis in software is the next step. Once synthesis is complete, another set of verifications and validations is performed using point tools. Azure can help teams accelerate this activity, like the verification done earlier. For example, project teams can launch multiple runs at the same time to optimize the step. Azure compute resources also provide massive scale for running a team's configuration management system for source control and project backups.

Back-end (physical) design

In back-end design, EDA software tools map the logical design to a specific foundry process. Steps include layout, static timing analysis (STA), physical verification, and tapeout based on GDSII. Testing includes design for test (DFT) practices and design validation using tools such as design rule check (DRC). In addition, a layout-versus-schematic (LVS) test in multiple iterations ensures that a design is accurately mapped and conforms to all the elements of the physical fabrication process.

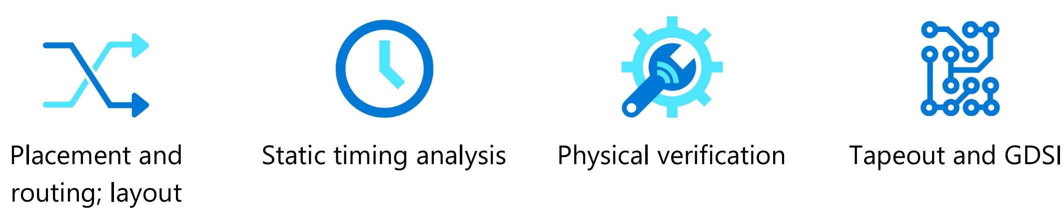

Figure 2. EDA software maps the logical design to the physical design and foundry process.

Many of these compute- and memory-intensive iterations require a very large number of cores and high-performance storage for huge datasets. Back-end design also includes a check on design for manufacturing (DFM), typically run in parallel batch processes to determine the best way to manufacture a silicon product.

Such workflows are ideally suited to run on Azure, where thousands of cores can be spun up to deliver a result in hours—not the days that would be required in an on-premises setup.

A different way to think about EDA workflows

When companies approach EDA infrastructure, they typically focus on maximizing the use rate of EDA licenses and making sure job runtimes are as fast as possible. To meet these goals, they invest in the highest clock speed. This thinking is used particularly in front-end design.

This thinking also applies in cloud-based infrastructures. In terms of overall performance, compute power is the biggest factor. However, the cloud poses other, equally important considerations, such as datacenter efficiency and workflow architecture. On Azure, the compute, storage, and scale components all affect performance. Our benchmarks show that storage in the cloud is a high-impact architectural component, as are scale technologies, which are essential to achieving the highest throughput. On Azure, EDA workflows can be optimized so that each toolset makes the best use of these components.

Azure versus on-premises: Efficiency

Datacenter efficiency considers the total cost of running a solution, where cloud infrastructures have several advantages over on-premises hardware. In an ideal world, you could run your solution on the very latest generation of CPU. Each generation can increase performance 10 to 20 percent, a significant advantage for EDA jobs. However, you can't guarantee that your jobs will run on the latest CPU in an on-premises datacenter, which most likely includes a range of CPU generations.

A cloud infrastructure built on the latest CPUs needs significantly fewer cores to run a job compared to a typical on-premises datacenter—even if it's only one or two CPU generations behind. In addition, even though the silicon industry's datacenter utilization targets tend to be very high, real-world data shows that it rarely exceeds 85 percent. A cloud-based infrastructure on Azure immediately provides a minimum 15 percent advantage in addition to cost savings, as you only pay for what you use.

Datacenter efficiency also considers the loss of productivity when engineers wait for jobs to run because the virtual machines (VMs) aren't available or don't match. This issue barely exists on Azure, where it's almost always possible to find the right VMs at the required scale.

Azure versus on-premises: Technology

Azure infrastructures have a technology edge over the typical on-premises datacenter. EDA workflows in the cloud can take advantage of massive storage with read and write throughput optimization in addition to scale-out features.

When we work with organizations that are moving EDA workloads to Azure, we use a granular approach:

- Evaluate the workflow and then provide each tool with its own optimized image and architecture.
- Fine-tune the storage specifications, networking configurations, and operating system version to achieve optimal performance.

In the past, it wasn't possible to optimize to this level of granularity. The sheer variety of ecosystem tools and investment choices made this approach impractical and ineffective. Now it is possible, thanks to the choices and flexibility that Azure provides.

We help our customers set up best-practice deployments while also supporting their rapid evolution. New, cloud-based technologies are published every six months, so we make sure they continue to run a deployment that features the latest configurations.

Azure also overprovisions hardware resources. EDA workloads may see a performance boost on Azure despite a parity on CPU clock rate. For example, overprovisioning the networking components at the rack level, coupled with SmartNIC cards on the motherboard, allows a singlehop CPU-to-CPU communication within a rack. This setup improves the throughput for EDA workloads that require heavy, cross-CPU communication. As another example, Azure holds back some CPU cores and RAM at the CPU, rack, and cluster levels and dedicates them for the virtualization function and network and storage I/O operations. This frees the virtual machine's CPU from such overhead.

Cloud decisions: Go hybrid or go all in

For many semiconductor businesses, the first major migration decision is whether to go hybrid or go all in. As Figure 1 shows, cloud resources can supplement an existing on-premises infrastructure, giving you a hybrid infrastructure capable of bursting workloads to the cloud when needed. Going *all in* means hosting the entire EDA workflow on Azure.

Each approach takes a unique framework, an architectural decision-making tree, and a plan for application and data migration. Azure has resources available to help every step of the way. The right approach depends on many factors, including where an organization is in its journey to the cloud.

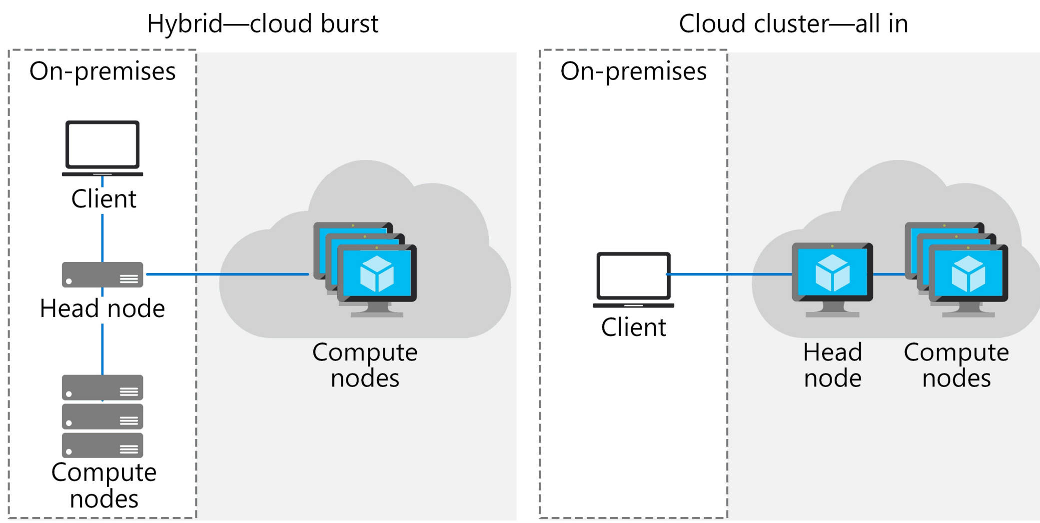

Figure 3. Hybrid versus all-in model with both the head and compute nodes in the cloud.

Burst to the cloud: The hybrid model

The recommended approach for organizations just starting their cloud journey is to go hybrid. This cloud-bursting approach uses cloud resources during peak hours or when on-premises resources are under stress.

The hybrid approach can help you support an existing infrastructure while adding compute and storage capacity on an on-demand basis. You can supplement the on-premises compute and storage resources used for legacy applications and workloads. When on-premises capacity is reached, you can run projects on Azure instead. Or you can offload specific workloads and their datasets to Azure. A team can burst verification jobs or an overnight regression on Azure, rapidly applying tens or many tens of thousands of compute cores to accommodate peak or unexpected demand.

In some hybrid use cases, the entire dataset for a process remains on an existing network-attached storage (NAS) on-premises. Only the active working set or needed tool binaries, design data, and libraries are burst into Azure before jobs are run. The working dataset can be manually uploaded to Azure or cached from on-premises to Azure to eliminate large data transfers on a per-workload basis. Most enterprise-class NFS data transfer solutions are also supported.

Going hybrid with specific workloads often improves monitoring, governance, and internal accounting. For example, some customers have preferred to move their most predictable lower priority jobs, such as daily or weekly regression tests, to the cloud while leaving the on-premises resources for higher priority jobs. Teams are free to optimize the compute environment on Azure for their workloads, datasets, and projects and to choose the best storage solution for their budget.

TSMC runs its IC design space, called VDE, on Azure. Similarly, many of our customers routinely run their silicon design and process development burst workloads on Azure.

Born in the cloud: The all-in model

An on-premises infrastructure for EDA workloads is costly to build, maintain, and expand. That’s why more organizations are choosing Azure infrastructure on demand to simplify operations, reduce costs, and speed solutions to market.

The all-in model works particularly well for startup businesses, which typically don’t have much (if any) on-premises infrastructure. They’re usually working on the latest applications and using limited resources to focus on and prioritize silicon design. A cloud-centric EDA environment and an all-in approach to Azure frees fast-moving teams from the limitations of a traditional infrastructure. Startups that adopt this model perform the full chip design within the cloud environment, taping out directly to the foundry.

Some semiconductor companies have adopted an all-in mindset. They choose to stop building or expanding their datacenters and take full advantage of the scale and flexibility of Azure. In this scenario, a company can choose to keep a percentage of its compute and storage requirements permanently on, and then scale in real time to match the demand while supporting EDA workflows on a high-performance infrastructure.

Microsoft works closely with the customers to optimize the cost of ownership of the compute infrastructure in the cloud, while maintaining the desired performance levels. With our understanding of finely tuned and highly utilized on-premises compute farms, we are able to drive new financial models and delivery mechanisms that make cloud solutions cost-beneficial for the silicon industry. Our team makes use of a spectrum of solutions to come up with a customized migration plan, orchestrated over time, to maximize the returns on the existing investments made to the on-premises infrastructure.

Many of the top semiconductor customers are making their journey through this process. In one such case, a large semiconductor company is considering doing its first advanced node project entirely in Azure, as doing it on-premises would require a serious upgrade to their infrastructure. In another example, TSMC recently used the Azure infrastructure to run a layout contest for university students in Taiwan to teach them how to combine process technology and chip design to optimize chip layouts and improve performance.[2]

---

[2] "TSMC Leads the Industry by Hosting the First ‘TSMC IC Layout Contest’ in the Cloud." Vivian Jiang. TSMC. March 12, 2020.

Aligning EDA ecosystems benefits the industry

Microsoft works in close partnership with the EDA vendors, foundries, infrastructure solution providers, and system integrators to align the entire silicon design ecosystem and help our customers make a successful migration to the cloud. For example, Mentor, a Siemens business, scaled its verification software to run across 10,000 cores on Azure. The library characterization runs that used to take days on-premises run in mere hours on Azure.[3] The company also set an industry record for scalability when it ran Calibre physical verification in a scale-out configuration of more than 4,000 CPUs on 5nm test chips.[4]

Similar work done by the Synopsys and Cadence teams has proven the scalability of their signoff timing and extraction solutions on Azure. For example, running Cadence solutions on Azure reduced semiconductor timing signoff schedules and cut costs in half.[5] Microsoft also worked with Synopsys and TSMC on a timing signoff flow for a multimillion gate design. The team tested the static timing analysis and the timing signoff software on Azure EDsv4-series virtual machines and saw significant throughput gains.[6]

In many of these cases, our work with the EDA vendors has resulted in improving the tool algorithms to prove the compatibility and scalability of these EDA tools in the cloud. This pre-work with the vendors allows our customers the best first experience when they try to burst in the cloud with these workloads.

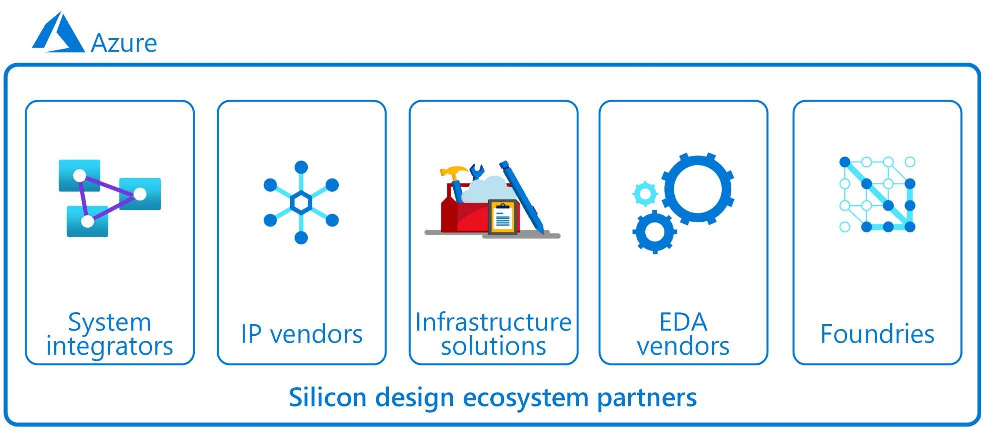

Figure 4. Microsoft partners with key industry players to enable the silicon design ecosystem.

Azure also supports the unique requirements of EDA workflows with options specifically designed for HPC workloads and the silicon industry.

- **EDA licensing.** Most EDA applications require a license from the vendor, whether the application is running on-premises or on Azure. A VM running EDA software in Azure can obtain license tokens from an existing on-premises license server across a network connection. A separate license server can also be provisioned on a VM in Azure. If dedicated

---

[3] "Mentor scales AMS cloud verification to 10,000 cores." Chris Edwards. Tech Design Forum. October 29, 2019.
[4] "Mentor sets an 'industry record' with Calibre physical verification." TechDecisions. June 7, 2019.
[5] "Cadence Collaborates with TSMC and Microsoft to Reduce Semiconductor Design Timing Signoff Schedules with the Cloud." Cadence press release. June 15, 2020.
[6] "Synopsys, TSMC and Microsoft Azure Deliver Highly Scalable Timing Signoff Flow in the Cloud." Synopsys press release. June 15, 2020.

EDA licenses are required to operate within Azure, you must procure them directly from the EDA vendors.

- **Support and training.** Microsoft provides training, engineering resources, and comprehensive support for every stage of cloud adoption. We also offer long-term support options.

Benefits of a globally available HPC platform

Regardless of the chosen approach or stage of implementation, semiconductor companies can expect Azure experts to help architect a best-fit solution with a plan for a seamless transition. Azure helps organizations meet current and emerging infrastructure needs for silicon design and development with a platform that provides:

- **Scalability.** Azure offers nearly unlimited scalability, as well as services to simplify the scaling process. Tools such as autoscaling and Azure CycleCloud programmatically allocate resources for optimal use. Azure Monitor provides infrastructure metrics and logs for most services in Azure. Organizations can also rapidly increase and decrease the number of cores needed, paying only for the resources used—an important benefit in the inherently cyclical business of silicon development.
- **Global presence.** Azure operates 54 global regions (and growing). This global presence offers semiconductor businesses the scale needed to bring applications closer to users around the world, facilitate design collaboration and data sharing, preserve data residency, and provide comprehensive compliance and resiliency options. Taking advantage of the Azure footprint, companies can reduce the cost, time, and complexity of operating a global semiconductor infrastructure.
- **Security.** Azure provides the security, privacy, and compliance protections used by 95 percent of Fortune 500 companies. Our multilayered approach to security starts at the foundation with the physical datacenters, infrastructure, and operations in Azure. Built-in controls and services in Azure extend protection to workloads and silicon intellectual property (IP) across identity, data, networking, and apps. Azure helps organizations maintain privacy and controls, meet compliance requirements, and ensure transparency. Microsoft has decades of experience building enterprise software and running some of the largest online services in the world. This experience is applied to continuously improve security-aware software development, operational management, and threat-mitigation practices that are essential to the strong protection of services and data.
- **Use models.** Azure fully supports burst, hybrid, and Azure-centric deployment, with efficient storage architectures for each scenario.
- **User authentication.** Most organizations already use an authentication service, such as LDAP, Active Directory, or NIS services. Azure provides hybrid identity solutions that span environments, creating a single user identity for authentication and authorization across resources, regardless of location. The Azure Active Directory Connect service provides integration between your on-premises directories and Azure Active Directory, providing a common identity for all your users.
- **Low-latency interactive jobs.** In addition to making a secure connection within the compute environment, interactive jobs also require a low-latency connection to ensure fast GUI response times when users make manual edits within a tool UI. With a large number of

regions, Azure has a datacenter within the vicinity of all major silicon design hubs across the world. Azure ExpressRoute is another way to speed network connections. A virtual desktop infrastructure (VDI) on Azure can replace specialized workstations and make EDA workflows accessible to more team members. Azure supports various virtual desktop solutions, including Windows Virtual Desktop, virtual network computing (VNC), NoMachine, NX-based systems, and Exceed TurboX (ETX), among others.

Azure Silicon Design Workbench

To address the need for a secure, scalable, performant silicon design environment for geo-distributed organizations and to support cross-organization collaboration, Microsoft is developing Azure Silicon Design Workbench. An intelligent, fully managed silicon design environment, Silicon Design Workbench provides multilayered security and access controls, along with the ability to monitor, scale, and optimize the compute capacity as needed.

Key benefits include:

- A scalable and resilient compute platform that is tuned for high-performance silicon design workloads known for their unique infrastructure requirements.
- A trusted environment for sharing and co-designing IP among various design teams, external design service partners, IP vendors, tool vendors, and other partners.
- Enhanced security and privacy using unique identity management and access control solutions, in addition to data partitioning and ownership policies, and the layer of data security features already offered by the Azure compute, storage, and networking infrastructure.
- The ease of platform as a service (PaaS) to automate the infrastructure build-up. Unlike high-touch IaaS solutions, Silicon Design Workbench decreases the time and resources invested in infrastructure and security management.
- Support for multiparty, multi-region collaboration across geographically dispersed engineering teams.
- An out-of-the-box solution for semiconductors and EDA customers interested in limiting their IT spend.

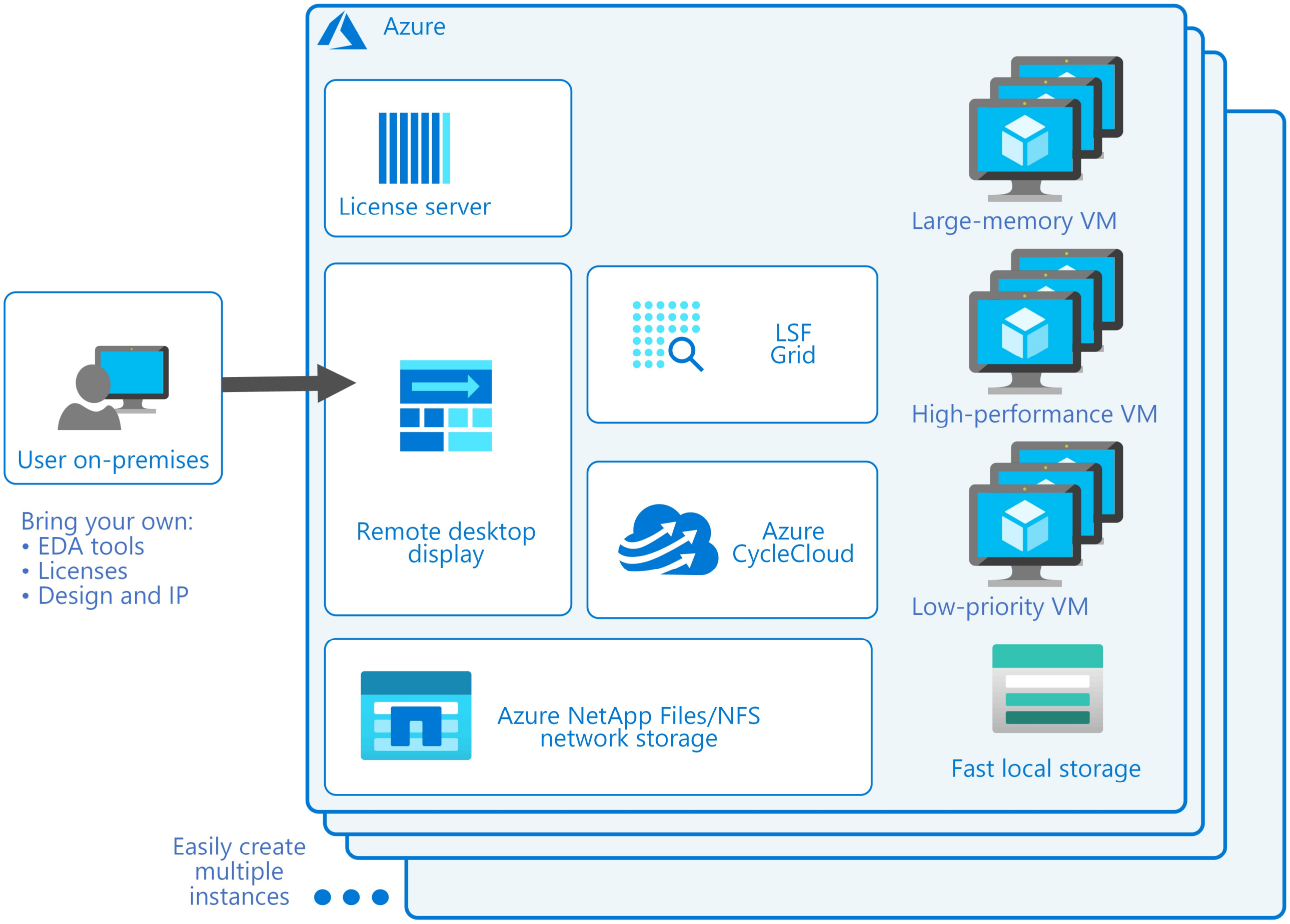

Figure 5. Azure Silicon Design Workbench is an automated, full-featured chip-design environment.

For more information about Silicon Design Workbench, contact SiliconWorkbench@microsoft.com.

Part 2: Azure architecture for EDA workloads

An Azure infrastructure for silicon design is optimized for compute- and memory-intensive applications, which are supported with high-performance file systems and efficient job scheduling to maximize throughput and performance of EDA software license investments.

Figure 6 shows a high-level architecture for EDA on Azure and introduces the compute, storage, networking, and orchestration components that are described in more detail in this section.

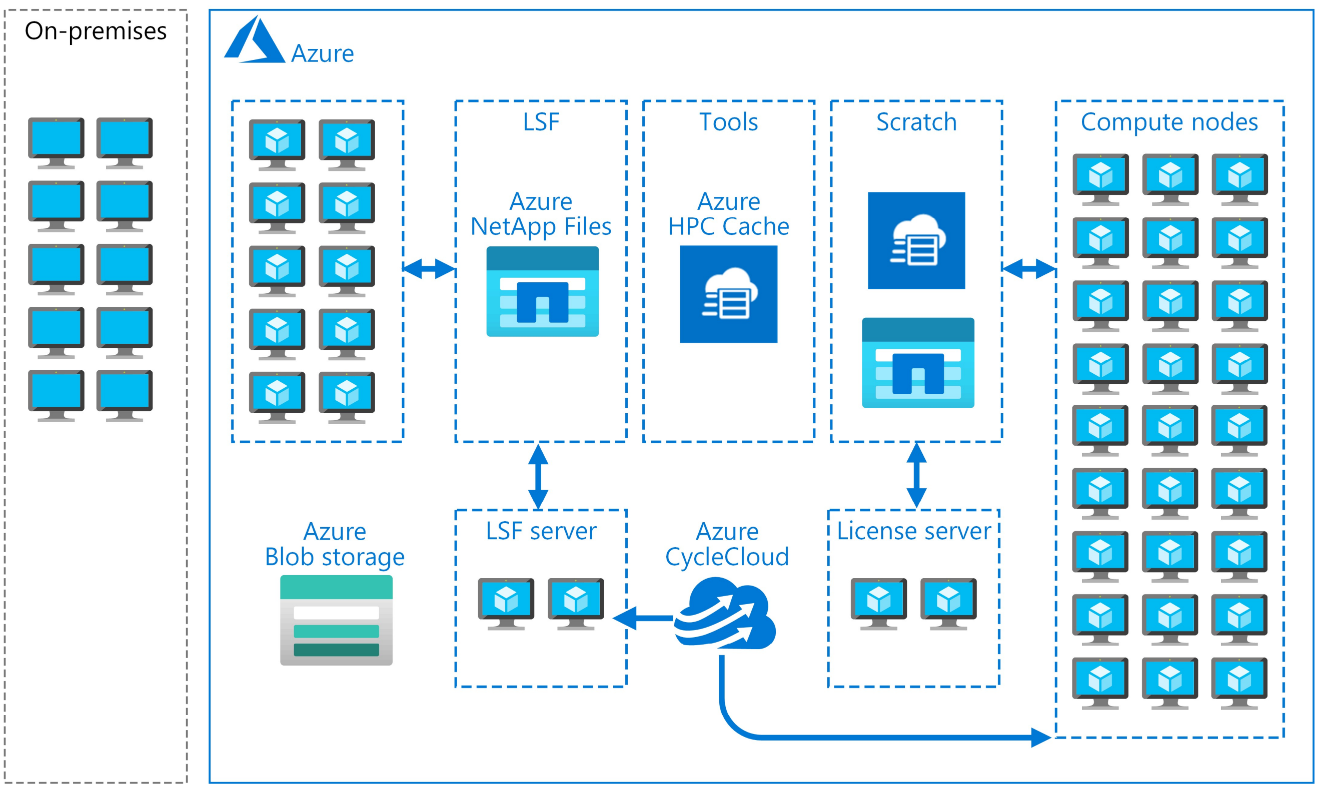

Figure 6. This high-level architecture supports EDA workloads on Azure.

This architecture includes the following components:

- **Azure compute.** Azure offers classes of virtual machines with a range of memory-to-core ratios that suit different workload requirements.
- **Azure NetApp Files.** This high-performance, metered file storage service enables you to migrate and run enterprise file applications without the need for code changes. Supported by Microsoft, Azure NetApp Files is built on the NetApp ONTAP storage OS.

- **Azure HPC Cache.** HPC Cache optimizes NFS latency and throughput of metadata and read operations to feed large, scale-out HPC clusters. HPC Cache supports up to 40 GiB per second of read throughput to a single file, provides microsecond responses to meta read requests, and scales out workload demands across millions of files and directories.
- **Azure Blob storage.** With exabytes of capacity and massive scalability, Blob storage stores from hundreds to billions of objects in hot, cool, or archive tiers, depending on how often data access is needed.
- **Azure CycleCloud.** This free tool is used to create, manage, operate, and optimize HPC clusters in Azure. For example, you can provision 50,000 compute cores in 20 minutes.
- **Networking.** The Azure Virtual Network infrastructure is based on software-defined networking (SDN) technology, where highly overprovisioned network resources can provide high bandwidth and low latency. An Azure ExpressRoute circuit is recommended to create a fast connection between Azure datacenters and the infrastructure on-premises.

In addition, Azure supports popular parallel virtual file systems, such as Lustre and BeeGFS, which are readily available in the Azure Marketplace.

Compute

Azure offers VM sizes that map to each stage of the silicon-development workflow. Multithreaded EDA workloads, for example, run well on the Fv2-series and the new Ddsv4 and Edsv4-series VMs based on the second-generation Intel Xeon Platinum 8272CL (Cascade Lake).[7] Fast clock speed and dedicated physical core (pCore) make the Edsv4-series, HC-series, and HBv2-series VMs well suited for single-threaded workloads, as well.

Azure supports Windows and Linux operating systems, including popular Linux distributions that are provided and maintained by our partners. For more information, see Endorsed Linux distributions on Azure.

General-purpose VMs

To host light infrastructure services, we recommend general-purpose VMs, such as Ddsv4, Dasv4, and Fv2. These sizes work well for license management, application UI servers, and other support tools with low-to-medium performance requirements.

Compute-optimized VMs

Compute-intensive EDA tools used in front-end design benefit from Edsv4, Easv4, Ddsv4, Dasv4, HBv2, and HC VMs. These series provide high-performance CPU and a high memory-to-core ratio and work well for specification, smaller block design, and simulation workflows.

For back-end applications used in placement and routing or for timing analysis, the Edsv4-series based on an Intel Cascade Lake CPU provides a suitably high memory-to-core ratio.

For physical verification, SPICE simulations, and other workloads that can run a large number of cores on a single VM, HBv2 is a good choice. It has 120 AMD EPYC 7742 processor cores, 4 GB of

[7] New general purpose and memory-optimized Azure Virtual Machines with Intel now available. Brenda Bell. Azure blog. June 15, 2020.

RAM per CPU core, and no multithreading.

Memory-optimized VMs

Memory-intensive workloads run well on Edsv4, Easv4, Mv1, and Mv2 series instances. For example, these VMs suit back-end implementation, physical and timing signoff verification, ERC checks, IR drop analysis, and physics and mathematical modeling, where memory requirements can reach up to 12 TB.

The latest memory-optimized VM is the Mv2-series, which offers the highest vCPU count (up to 416) and the most memory (up to 11.4 TiB) of any VM in Azure.

Storage

Three key storage components work together to provide high performance for EDA workflows—Azure NetApp Files, Azure HPC Cache, and Azure Blob storage.

Azure NetApp Files

Azure NetApp Files makes it easy to migrate and run demanding EDA file workloads on Azure. Powered by NetApp technology, Azure NetApp Files is a fully managed cloud service from Microsoft that provides the same or better performance than the version organizations run on-premises. In addition, deploying the service takes minutes—not the days or weeks it takes to set up NetApp in a datacenter.

Azure NetApp Files supports three service levels: Standard, Premium, and Ultra. Each level provides different network throughput, as the following table shows.

Table 1. Service tiers and throughput in NetApp Files[8]

Tier ▸	Standard	Premium	Ultra
Throughput ▸	16 MiB/s per 1 TiB	64 MiB/s per 1 TiB	128 MiB/s per 1 TiB
	Up to 1,000 IOPS/TiB	Up to 4,000 IOPS/TiB	Up to 8,000 IOPS/TiB

NetApp Files is built on the Data ONTAP storage operating system. Microsoft and NetApp have partnered for decades to build increasingly powerful and flexible replication and migration features. For example, NetApp Files encrypts data at rest in compliance with FIPS 140-2 and supports role-based access control (RBAC) and network access control lists (ACLs).

NetApp Files also delivers enterprise-grade security. The service is physically hosted inside Azure datacenters and operates within Azure security boundaries. All interactions with critical systems strictly follow just-in-time and just-enough-access rules—users are granted the minimum level of privilege for the least amount of time to perform critical actions. The only access to production systems is from specially secured workstations through multifactor authentication. All resources that deliver NetApp Files functionality require security software, and all code down to the firmware is scanned.

[8] For information about capacity-per-hour costs, see Azure NetApp Files Pricing.

Figure 7 illustrates a fully managed NetApp Files service that is deployed in an organization's Azure subscription.

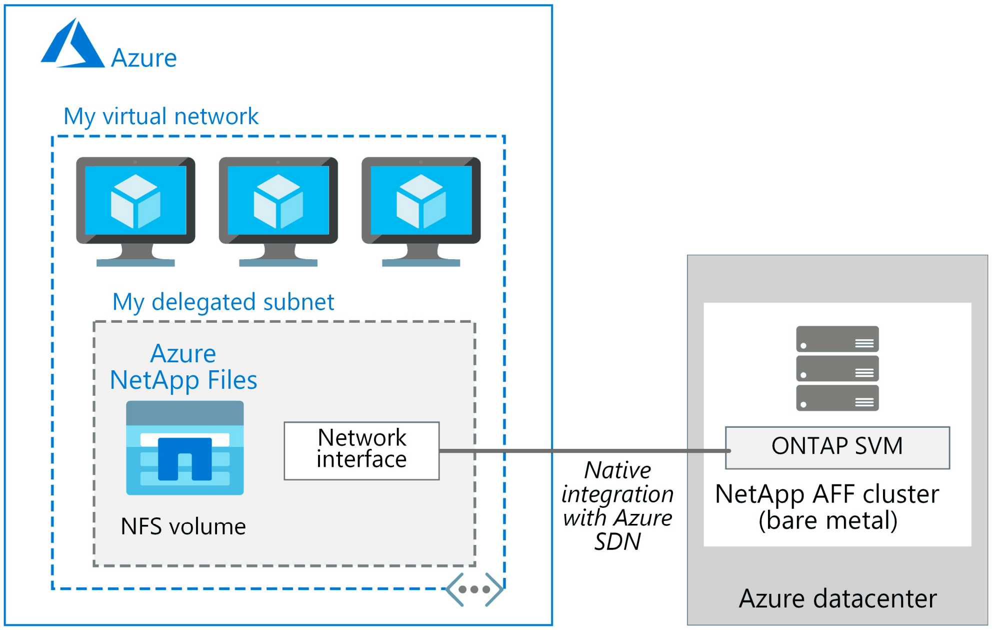

Figure 7. NetApp Files is a fully managed service that is deployed to an Azure subscription.

For more information, see Benefits of using Azure NetApp Files for electronic design automation in the Azure documentation.

Azure HPC Cache

EDA workflows are tough on storage subsystems. To minimize the I/O wait time, these workflows need low-latency and high-throughput responses to the high-concurrency NFS requests made by core compute jobs. Azure HPC Cache is the infrastructure service that seamlessly serves NFS data to compute grids as they scale from hundreds to tens of thousands of cores.

As a managed NFS caching service, HPC Cache simplifies the tasks of managing data and scaling the compute grid. The service gives each grid compute node access, using NFS, to the binaries, tools, libraries, and staged design files it needs. Semiconductor companies are using HPC Cache in production EDA clusters to support 80,000 cores and up and to reliably serve millions of file and directory metadata IOPS with an NFS read throughput of tens of gigabytes per second.

Key benefits include:

- High-performance that reduces latency for read-heavy EDA workloads, such as the tools and projects repositories, with up to 8 GB per-second throughput.
- Dynamic scalability that meets changing compute demands across the phases in the EDA process.
- A namespace that brings together multiple storage targets (local or remote NFS, or Blob). Clients can navigate the storage targets using a single NFS mount point in HPC Cache, and the file system contents are cached locally.
- Easy monitoring and management through the Azure portal or through the APIs provided in the SDK.

In addition, HPC Cache works well in cloud-bursting scenarios for organizations that need extra resources during peak hours or additional capacity to supplement an on-premises datacenter. The physical cluster can be used to cache data to Azure, making it easier to shift NFS-based EDA workloads to Azure. HPC Cache also supports existing large-scale workflows that require NAS filer data residing on-premises. (In this scenario, an ExpressRoute connection is recommended.)

Azure Blob storage

Azure Blob storage is a massively scalable, cost-effective alternative to on-premises solutions for cold data storage. Optimized for unstructured data, Blob storage provides exabytes of capacity and REST-based object storage for long-term archives.

Networking

Network quality and throughput make a significant impact on EDA job runtimes. Azure provides built-in and customized options for fast, scalable, and secure connectivity between your datacenter and the global Azure regions. Microsoft investments in private optical-fiber capacity and undersea cabling enable low-latency access globally, as well as peering between regions for wide-footprint companies. Microsoft owns and runs one of the largest WAN backbones in the world.

For large transfers of data, Azure offers online, offline, and several supported partner solutions. For example, during a single extraction and timing signoff job, a large volume of data is exchanged between the VMs and flows to and from the shared NFS storage.

For the most demanding, memory-intensive EDA workloads, Azure offers accelerated networking, which enables single root I/O virtualization (SR-IOV) to VMs that support this feature, such as the Esv4-series. Accelerated networking makes it possible to move much of the Azure software-defined networking stack off the CPUs and onto SmartNICs. An application of field-programmable gate arrays (FPGAs), SmartNICs enhance NICs with reconfigurable logic. Applications can reclaim compute cycles, reducing latency, jitter, and the load on the VMs.

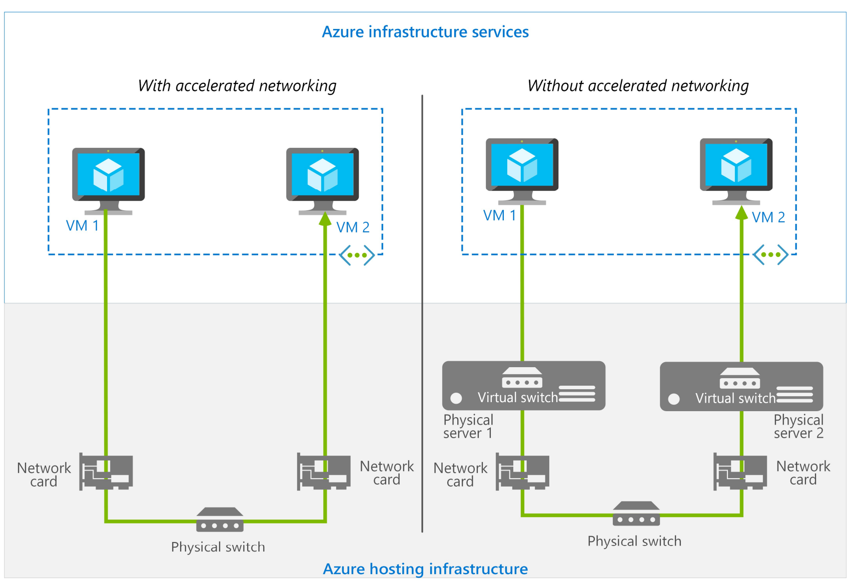

Figure 8. Accelerated networking reduces the number of hops and improves performance.

Azure ExpressRoute

The Azure ExpressRoute network service creates private connections between Azure datacenters and an infrastructure on-premises or in a colocation environment. Because ExpressRoute connections do not traverse the public internet, they offer more reliability, faster speeds, and lower latencies than typical internet connections. Predictable latency contributes to more predictable performance for EDA workloads. In some cases, using ExpressRoute connections to transfer data between on-premises systems and Azure can also produce significant cost benefits.

An ExpressRoute connection to Azure can be established from an ExpressRoute location or directly from an existing WAN network (typically MPLS VPN) from a network service provider.

For more information, see [ExpressRoute connectivity models](#).

Remote desktop solution

Some EDA applications are graphically demanding software, and users need to access EDA workloads' desktop environments, such as GNU Network Object Model Environment (GNOME) or Kool Desktop Environment (KDE), remotely.

As robust security is a must to protect silicon IP, the remote desktop access must be secured, for example, by blocking data transfer through the sessions. Another key requirement is a high level of access control, such as using Azure Multi-Factor Authentication and assigning access permissions to users of EDA workloads.

To centralize secure remote access and simplify monitoring and logging, organizations typically deploy a dedicated remote desktop solution. For example, OpenText Exceed TurboX provides remote desktops for securely controlling EDA assets on Azure.

Orchestration with Azure CycleCloud

EDA workloads can consume anywhere from tens to thousands of cores in a single run. Multiple parallel runs, varying machine configurations, multiple project needs, last-minute design changes, and other considerations can add to the complexity of right-sizing an on-premises IT infrastructure.

Azure CycleCloud is an orchestration service used to create, manage, operate, and optimize HPC and big compute clusters of any scale. Available at no charge to Azure customers, this end-to-end tool helps IT administrators and developers create, manage, use, and optimize dynamic clustered-compute environments. With Azure CycleCloud, users can choose how to deploy on-premises and on Azure and dynamically grow or shrink compute capacity as needed.

Azure CycleCloud supports file servers, select parallel file systems, and popular job schedulers, such as Oracle Grid Engine, Slurm, and Altair PBS Professional, to provision compute resources based on the requirements of the jobs in the queue. In addition, IBM Spectrum LSF officially supports and provides a connector for Azure CycleCloud. For details, see Submitting jobs to launch instances from Azure CycleCloud in the IBM Knowledge Center.

Key CycleCloud benefits include:

- Efficient, dynamic scaling of EDA workloads based on work queues, from one to thousands of instances. No application rewrites are required to enable existing EDA workflows.

- Support for almost any job scheduler, application stack, or cluster configuration. For example, CycleCloud works directly with EDA job schedulers to autoscale based on job status.
- Easy EDA cluster setup based on roles and groups of nodes with corresponding setup and initialization flows.
- Management options that make it easy to track and monitor total CPU hours and define the maximum number of CPUs to acquire.
- Policy and governance features for managing costs, users, and access.

CycleCloud cluster and resource deployment

Azure CycleCloud simplifies the task of defining and provisioning clusters for EDA workloads. It offers expressive and parameterizable cluster templates for deploying Azure infrastructure, including networking and storage components, in addition to compute resources, such as VMs, GPUs, and managed disks.

Using cluster templates, organizations ensure consistent, repeatable, easily extensible cluster configurations. Templates represent the best practices for EDA cluster deployments, including the resource types and configurations best suited for EDA use cases.

Cluster administrators have the option to use Azure CycleCloud to abstract the cloud from users, providing a controlled and managed interface to scalable compute. CycleCloud also provides self-service capabilities, along with features for managing access, security, and costs.

CycleCloud policy and governance services

CycleCloud provides orchestration governance and policy enforcement that help organizations manage an Active Directory integration, control and report on costs, and support error-handling. This includes policies such as:

- Who has access to specific compute nodes, and how much?
- The versions of workflows to execute, and which infrastructure to tailor for a specific workload.
- Cost and budget limits per cluster, per user, and other options.

For example, cluster administrators can set policies that define a service-level agreement instead of managing compute resources manually according to per-user or other limits. The policy and governance services offer unique flexibility for managing compute resources in a secure, cost-effective way.

CycleCloud for storage orchestration

CycleCloud can also orchestrate the deployment of various storage options, including Azure NetApp Files and HPC Cache. Organizations can also define and manage storage components separately or as part of the compute clusters. Administrators and users can manage the data life cycle apart from compute, helping organizations to optimize deployments for cost, access, and performance as appropriate.

Azure DevOps

For the silicon industry, DevOps practices are essential for shortening the time required to move a design from the planning stage to production and delivery. These practices replace the siloed development and operations efforts of the past to create multidisciplinary teams that work together with shared and efficient practices and tools.

Azure DevOps provides a set of fully managed services that help teams plan work, collaborate on code development, and build and deploy applications. These services include:

- Azure Boards for planning, tracking, and discussing work.
- Azure Repos for unlimited, cloud-hosted, private Git repos. Azure Repos connects from any IDE, editor, or Git client to provide collaborative pull requests, advanced file management, and more, for teams and projects of any size.
- Azure Pipelines for continuous integration (CI) and continuous deployment (CD). Pipelines works with any language and platform, connects to GitHub or any Git provider, and deploys continuously to on-premises, hybrid, and cloud environments.
- Azure Test Plans for test management. It includes a Test & Feedback extension that makes it easy for teams to perform exploratory testing and provide feedback.
- Azure Artifacts for creating host and share packages that add artifacts to CI/CD pipelines. With consistent administrative and access control, Artifacts makes it easy to share code across small and large groups and supports Maven, npm, and NuGet.

Partner solutions for revision control

Azure supports other popular revision control solutions for EDA workflows:

- Git is an open-source, distributed revision control system with a small footprint and fast performance.
- Perforce Helix Core version control software supports EDA and semiconductor design with standard process flows. Perforce works closely with customers to adapt their unique process and development workflows and specific performance requirements to Azure.
- IBM Rational ClearQuest and IBM Rational ClearCase work together to provide a unified change management solution for Azure.
- Subversion with migration including history.
- Concurrent versions system.

Part 3: Next steps

Azure offers a strategy for today's cloud migration and a platform for developing tomorrow's innovations. Microsoft has helped some of the world's largest semiconductor companies to power the most infrastructure-demanding EDA workflows on Azure. These companies can add and remove resources as demand changes and pay only for what they need, when they need it. Azure also provides robust security that helps protect their silicon intellectual property. Trusted Azure solutions give companies the flexibility to choose the cloud model that best fits their needs—burst to the cloud in high-demand times or move entire workloads into Azure.

Six steps to get started

Many semiconductor IT organizations struggle with the decisions of what to move when—deliberations that can be complex and time consuming. We recommend a six-step process for getting started on Azure:

1. Catalog your software and workloads.
2. Categorize performance and workloads.
3. Define success criteria for moving to or starting a workflow in Azure.
4. Architect core infrastructure components for cloud integration.
5. Get the skills you need for development.
6. Develop a cloud production support model and retool for adoption and change management.

Contact your Microsoft account team for more information about Azure semiconductor solutions.

Resources

Microsoft offers a wealth of resources to assist semiconductor companies with their journeys to the cloud. Azure can simplify and streamline that process at every stage, from early assessment of requirements and available services to workload optimization for all-in cloud deployments.

For example, we offer cloud workshops and educational services for IT staff and decision-makers so they can learn about Azure and prepare their teams to evaluate and adopt public-cloud solutions. Companies that are further along in the cloud journey that need solutions now can work with us and start benefiting from the reliability, scalability, and security of the Azure infrastructure. We partner with EDA customers at every step to help identify the best path forward and then architect a solution and help ensure a successful deployment.

- For more information about our position on smart manufacturing in the silicon industry, see Azure HPC for silicon.
- For a high-level look at HPC architectures on Azure, see HPC on Azure Overview.

In the news

"Cadence Collaborates with TSMC and Microsoft to Reduce Semiconductor Design Timing Signoff Schedules with the Cloud." Cadence press release. June 15, 2020.

"Synopsys, TSMC and Microsoft Azure Deliver Highly Scalable Timing Signoff Flow in the Cloud." Synopsys press release. June 15, 2020.

"New general purpose and memory-optimized Azure Virtual Machines with Intel now available." Brenda Bell. Azure blog. June 15, 2020

"How cloud computing is now delivering efficiencies for IC design." Omar El-Sewefy. Tech Design Forum. May 26, 2020.

"Cloud Accelerated Idea To Silicon." Chris Lattner. SiFive blog. March 25, 2020.

"SiFive Selects Synopsys Fusion Design Platform and Verification Continuum Platform to Enable Rapid SoC Design." Synopsys press release. March 25, 2020.

"Chip Design and the Azure Cloud: An Azure NetApp Files Story." Chad Morgenstern. NetApp blog. Mar 17, 2020.

"TSMC Leads the Industry by Hosting the First 'TSMC IC Layout Contest' in the Cloud." Vivian Jiang. TSMC. March 12, 2020.

"Mentor Adds Circuit Simulators to the Cloud using Azure." Daniel Payne. SemiWiki.com. November 8, 2019.

"Mentor's analog/RF/mixed-signal verification tools scale to 10,000 cores on Microsoft Azure." Mentor press release. October 29, 2019.

"TSMC in the Cloud Update #56thDAC 2019." Daniel Nenni. SemiWiki.com. June 13, 2019.

"Mentor and AMD verify massive Radeon Instinct Vega20 IC design on AMD EPYC in ~10 hours with ecosystem partners Microsoft Azure and TSMC." Mentor press release. May 30, 2019.

"TSMC Strengthens OIP Cloud Alliance with New Partner and New Solution Enablement." TSMC press release. April 26, 2019.

"Cadence Extends Cloud Leadership With New CloudBurst Platform for Hybrid Cloud Environments." Cadence Design Systems press release. April 4, 2019.

"Mellanox Uses Univa to Extend Silicon Design HPC Operation to Azure." Doug Black. HPCwire. December 11, 2018

"TSMC Announces OIP Ecosystem Enabled in the Cloud." TSMC press release. October 3, 2018.

"Synopsys Announces Availability of TSMC-certified IC Design Environment in the Cloud." Synopsys press release. October 3, 2018.

"Cadence Expands its Cloud Portfolio with Delivery of TSMC OIP Virtual Design Environment." Cadence press release. October 3, 2018.

"Cadence and Microsoft Collaborate to Facilitate Semiconductor and System Design on the Microsoft Azure Cloud Platform." Cadence press release. June 25, 2018.