Azure for the semiconductor industry

Silicon-design workloads with electronic design automation software

Azure Silicon Industry Team
AzureForSilicon@microsoft.com

November 2021

Contents

List of figures

Authored by the Azure Silicon Industry Team: Jon Austin, Bernie Behn, Andy Chan, Bella Chan, Preeth Chengappa, Giancarlo DePasquale, Mujtaba Hamid, Sooyoung Moon, Richard Paw, Michael Requa, Steve Roach, Raymond Tsai, Prashant Varshney, Gretchen Weaver

Special thanks to our partners at NetApp, Perforce, and Pure Storage.

Executive summary

This document demonstrates why Azure is the optimal cloud platform for silicon design; our focuses on tools and technologies that are already being successfully used by some of our largest customers in the semiconductor industry and have been validated through collaboration with industry partners.

Semiconductors enable the modern world. They're the critical component in everything from communications to IoT (Internet of Things) devices, supercomputers to dishwashers. They enable cars to be more fuel efficient than ever and help researchers with drug discovery. Growing demand for more capable semiconductor devices forces an ever-increasing rate of innovation from the industry.

Shorter product cycles, new process geometries, and growing design complexity require more effective design flows and more design resources to meet schedule demands. Each new process geometry introduces an order of magnitude more design rules which engineers and design tools must account for. Increasing performance needs require increased integration of features leading to larger SoCs (System-on- a-Chip), which also require more verification and validation. New technologies, such as 3D-IC and other innovations to increase the functionality per square millimeter of silicon, introduce the need for new innovations in EDA tools and even more, and increasingly, complex validation. The need to accommodate these factors has led to the exponential growth of compute resource requirements by the semiconductor industry.

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, EDA vendors and users alike to develop finely tuned solutions that run on Azure high-performance computing (HPC) infrastructures. Our cloud-based solutions include optimized reference architectures covering for each EDA tool. As such, semiconductor companies are leveraging Azure solutions to improve the quality of their designs, develop innovative solutions, and accelerate 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 time windows to deliver greater functionality, higher quality, and more customizations. Product teams can trade-off budget vs. verification coverage for a given schedule window.

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.

Part 1: Silicon design in Azure

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 and saleable 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, OSATs (Outside Semiconductor Assembly and Test), and system integrators (implementation consultants).
- 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 tape-out for the back-end design—has a dedicated set of EDA tools. Azure services and developer 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 with some block-level simulation and debug. This workflow benefits from Azure DevOps tools and services, which improve the design cycle by fostering team collaboration and agile practices.

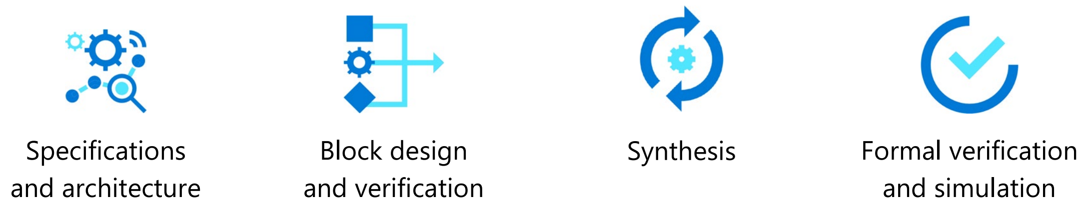


Figure 1. Steps in the front-end design phase use software simulation.

As teams develop the design blocks, the design elements are verified for correctness by running functional simulation. RTL (Register Transfer Level) simulation enables designers to confirm that the design will logically function as expected. To minimize time spent in simulation, these various design elements are simulated in parallel, across dozens to thousands of systems, depending on the size of the design and resources available. Azure, with its near-infinite compute resources, can accelerate the verification process by allowing more jobs to run in parallel, thereby shortening the total turn-around-time.

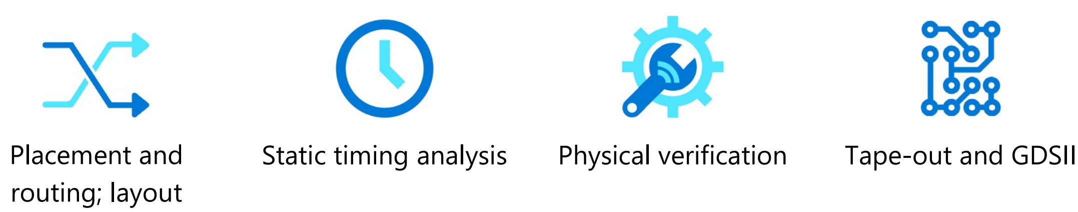


Figure 2. EDA software maps the logical design to the physical design and foundry process.

Many of these back-end workloads have very large compute and memory requirements. These workloads can scale to high numbers of cores and rely on high-performance storage for access to huge datasets. For many product teams, even those in large companies, the availability of such large core count and memory systems are a challenge. Even if the systems are in the data center, the demands for access to these few systems presents a challenge in aligning availability of access to run large back-end workloads and can often restrict the scale of these workloads, prolonging the turn-around-time.

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 data center 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

Data center 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 data center, 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 data center—even if it's only one or two CPU generations behind. In addition, even though the silicon industry's data center 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.

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

Azure versus on-premises: Technology

Azure infrastructures have a technology edge over the typical on-premises data center. 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 can 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 technology on the motherboard, allows a single-hop 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.

Azure versus on-premises: Flexibility

Many of the limitations placed on the design flow and methodology for existing chip design are a function of the limitations of current semiconductor design infrastructure. For example, the limitation that all tools in each design flow all must run on the same version of a given operating system and on the same computational technology is a function of the limitation of the compute available in each company's data centers. This limits the versions of tools available to a given product team. A product team starting a new design may be limited to older versions of tools because an existing design in flight is running on a previous version of the operating system and the need to run on the same systems can limit options available to the new product team. On Azure, each design phase (workload) for each project runs on its own bespoke cluster of VMs. This allows each workload to run on the best combination of operating system and technology the product team chooses since each workload is independent of the others from an infrastructure perspective. Product teams can choose the tool version they want and even leverage GPUs and AI technologies that are available on Azure, but not available in-house.

For highly parallelizable jobs, such as simulation and library characterization, product teams can choose to scale these workloads out horizontally in order to reduce the total wall clock time required to process the verification suite without being limited by how many systems are available in-house. Product teams can choose to reduce the amount of turn-around-time for the verification queue by purchasing access to additional licenses and VMs if schedules demand it. They can also decide to increase coverage and, thereby reduce their tape-out risk, by running additional verification jobs at the same time.

For large, memory intensive back-end workloads, the availability of systems with memory sizes in the multiple TB of RAM means being able to right-size the system to the workload. Design teams can choose the optimal number of CPU, amount of RAM, number of systems, etc., for their workload, on their schedule, without having to wait for system availability to align. With systems available that have up to 12TB of RAM, full-chip verification workloads that were previously impossible to run on-premises are now possible on Azure.

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 3 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.

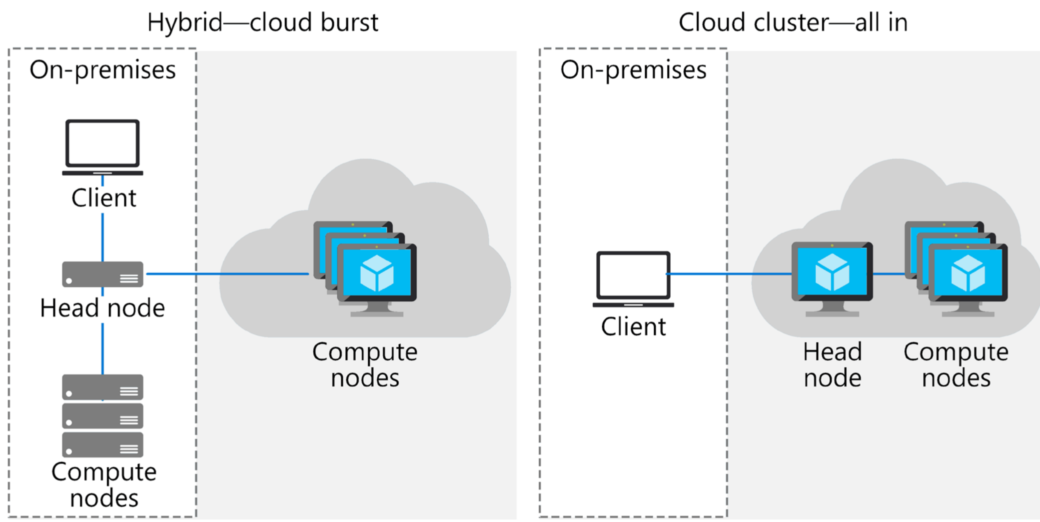


Figure 3. Hybrid versus all-in model with both the head and compute nodes in the cloud.

Burst to the cloud: The hybrid model

The most common approach for organizations just starting their cloud journey is to go hybrid. Most EDA customers start by either targeting highly parallelizable workloads such as simulation and library characterization, or a workload that requires a system with large memory that they may be in short supply of, such as STA, DRC, or IR Drop sign-off. 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. When on-premises capacity is reached, you can expand workloads onto Azure instead. Customers can choose to 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 allowing many workloads to share the same data 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) pre-existing on-premises infrastructure. They're usually working on the latest product and using limited resources to focus on and prioritize silicon design. A cloud-centric EDA environment and an all-in approach with 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. Startups are also able to focus more of their limited resources on maximizing the amount of available engineering and compute power without having to spend money upfront on infrastructure that depreciates over years, or the real-estate required to house them. Resource constrained start-ups also do not have to divert a valuable engineer's time or hire a contractor to handle more mundane infrastructure IT chores such as hardware installation and maintenance, procurement, provisioning, etc.

Some semiconductor companies have adopted an all-in mindset. They choose to stop building or expanding their data centers 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 can 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.[1]

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

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

customers make a successful migration to the cloud. For example, Siemens EDA, formerly Mentor Graphics, scaled its library characterization software to run across 10,000 cores on Azure[2]. Jobs that could take days on-premises now 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]

Azure also worked with Ansys to test Redhawk-SC on a 50 billion electrical node full-chip run using the FX VM Family.[7] This configuration demonstrated scalability over 240 CPUs. Ansys and Azure also collaborated to solve an adaptively converged mesh (15 passes) for an entire RFIC (5.5 x 5.5mm) at 5GHz in under 30 hours. This model has 64 ports distributed across the entire IC, representing the most challenging conditions for the adaptive refinement process to solve.[8]

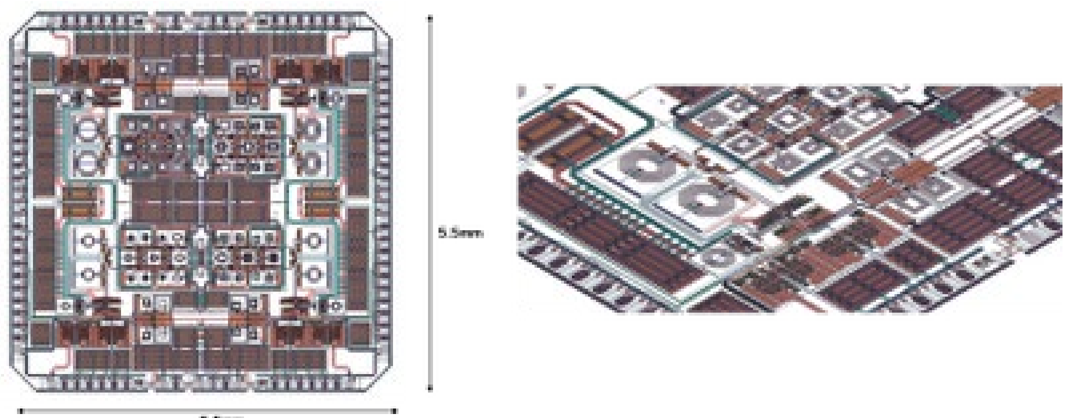


Figure 4. Top and Angled view of Integrated Circuit in HFSS.

- Compute cores used: 704 cores (Intel Xeon Platinum 8168, Azure “HC44” VM)
- Initial Mesh solve time: 1h54m

---

[2] “Mentor’s analog/RF/mixed-signal verification tools scale to 10,000 cores on Microsoft Azure”, Siemens newsroom, Oct 29th, 2019
[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.
[7] How Azure FX VM Makes Ansys RedHawk-SC™ Run Faster the Less You Spend “Microsoft tech community release June 22nd, 2021
[8] It Was Impossible – Until Now. Computational Electromagnetic Breakthrough on Azure. “Microsoft tech community release November 16, 2020[9] New general purpose and memory-optimized Azure Virtual Machines with Intel now available. Brenda Bell. Azure blog. June 15, 2020.

- Final RAM used: 2.6TB
- Final mesh size: 23.5M Tetrahedron and 93M unknowns
- Final mesh point solution time (5GHz frequency point solved after adaptive pass 15) = 2h19m
- Total Adaptive Mesh Time: 29h47m

In many of these cases, our work with the EDA vendors resulted in improvements to 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.

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. However, due to latency between the license server and the VM on cloud, some tools may suffer performance degradation due to latency in license response. To address this issue, a separate license server can also be provisioned on a VM in Azure. If dedicated 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 EDA 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, provide comprehensive compliance and resiliency options, as well as disaster recovery and business continuity. 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 data centers, 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 many regions, Azure has a data center 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.
- Facilitate the frictionless creation of a secure collaboration chamber that allows customers full control of what data, tools, licenses, and resources the chamber has access to, as well as which external user gets access, when they can access the environment, and how much access they have.

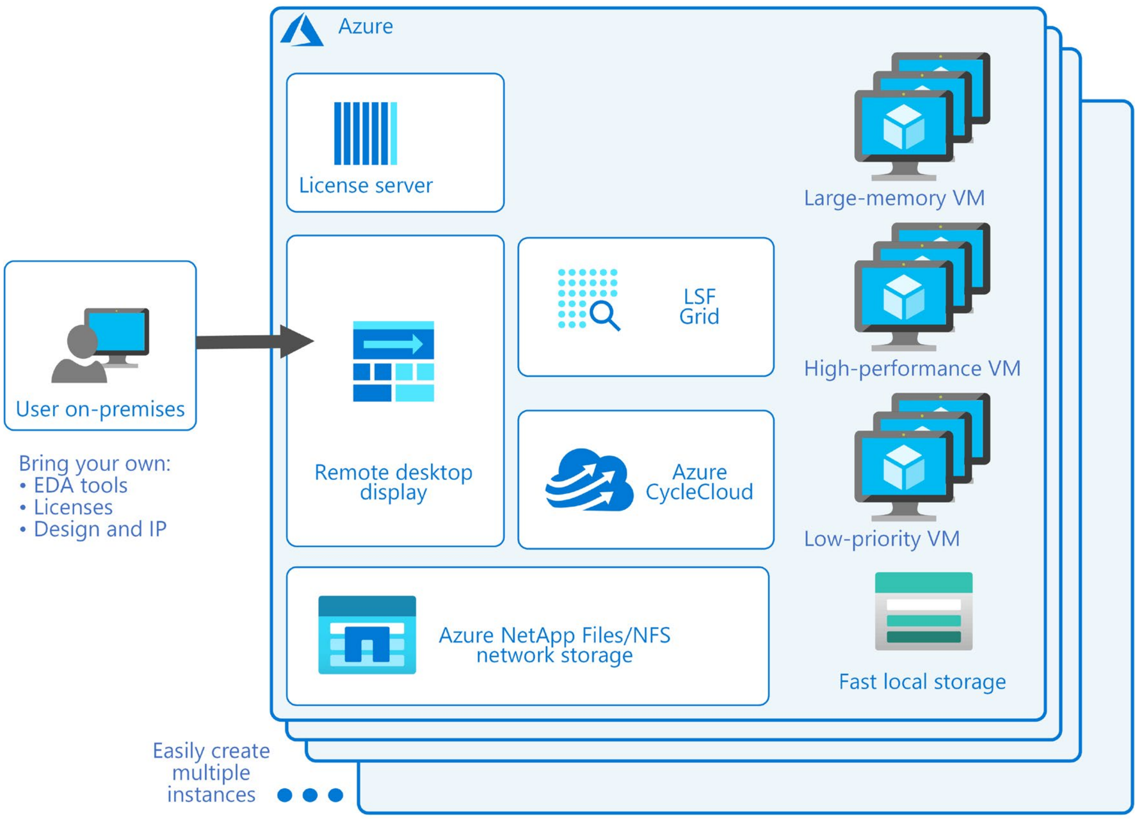


Figure 5. Azure Silicon Design Workbench is an automated, full-featured chip-design environment.

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.

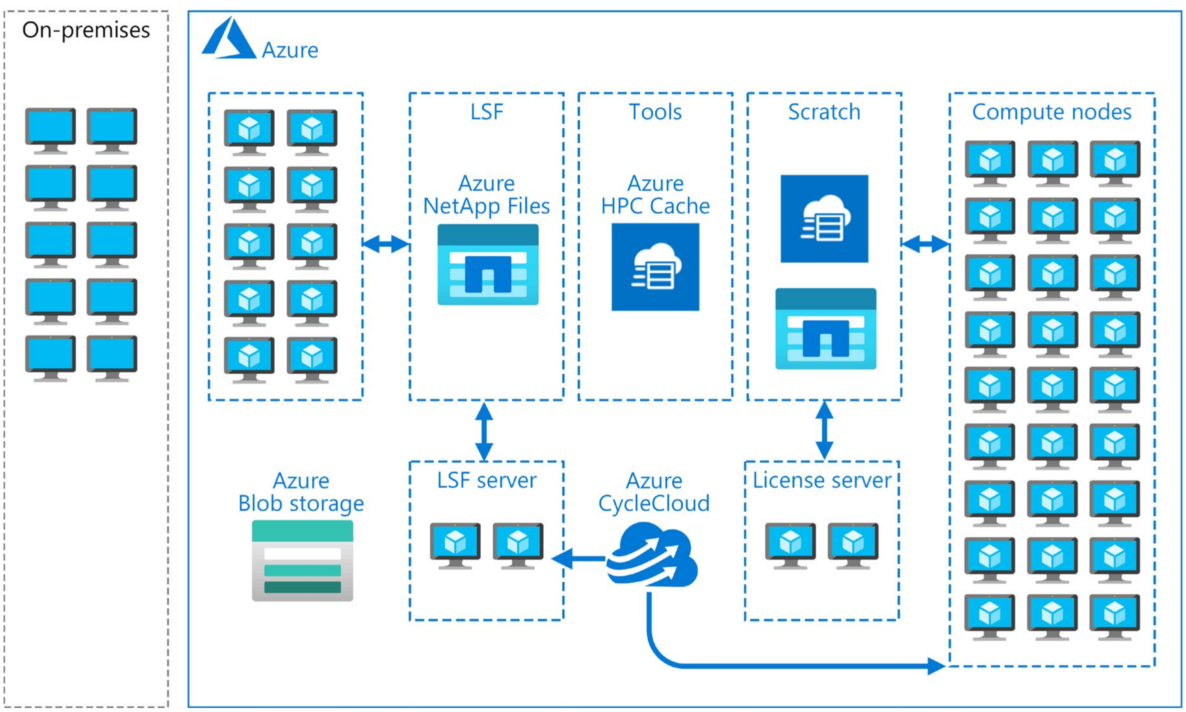


Figure 6. This high-level architecture supports EDA workloads on Azure.

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 file-based EDA tools without the need for code changes. Supported by Microsoft, Azure NetApp Files is built on the NetApp ONTAP storage OS providing EDA customers with an interface and features they're already familiar with.
- **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 20 GiB per second of read throughput to a single file, provides microsecond responses to metadata read requests, and scales out workload demanding millions of files and directories, reducing the contention on high demand read volumes and network bottlenecks.

- **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 Blob storage provides a more cost-effective way to store decades of historical design data while maintaining instant accessibility to that data.
- **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 data centers 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 Ddsv5 and Edsv5-series VMs based on the second-generation Intel Xeon Platinum 8370C (Ice Lake) and later.[9] Fast clock speed and dedicated physical core (pCore) make the Edsv5-series, HC-series, and HBv3-series VMs well suited for single-threaded workloads, as well.

Azure supports Windows and Linux operating systems, including RHEL, SLES, and CentOS which are commonly used in EDA workloads. For more information on these and other supported distributions of Linux, see Endorsed Linux distributions on Azure.

General-purpose VMs

To host light infrastructure services, we recommend general-purpose VMs, such as Ddsv5, Dasv5 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 Edsv5, Easv5, Ddsv5, Dasv5, HBv3, FXv1, 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.

Memory-optimized VMs

Memory-intensive workloads run well on FXv1, Edsv5, Easv5, Mv1, and Mv2 series instances. For example, these VMs suit back-end implementation, physical and timing signoff verification, ERC

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

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.

EDA focused VMs

Two recently introduced VMs were developed with EDA in mind. Learnings from the close collaboration between semiconductor customers, EDA partners, and Microsoft guided the development of the FXv1 and HBv3 VMs.

FXv1

The Fxv1 features the Intel Xeon Gold 6246R (Cascade Lake) processor with all core turbo of 4.0GHz sustained. It has up to 24 physical cores, or 48 hyper-threaded virtual cores. That gives the FXv1 42GB/physical core of memory. It also comes equipped with 2TB of local NVMe storage that can be used for local /scratch space or as virtual memory for extremely large designs.

With almost a Terabyte of RAM, the FXv1 can run full chip back-end workloads for most designs today and in the near future. 42GB/physical core combined with the 4.0 GHz processor makes the FX a good candidate for large scale simulation workloads.

HBv3

HBv3 VMs feature up to 120 AMD EPYC™ 7003-series (Milan) CPU cores, 448 GB of RAM, and no hyperthreading. HBv3-series VMs also provide 350 GB/sec of memory bandwidth, up to 32 MB of L3 cache per core, up to 7 GB/s of block device SSD performance, and clock frequencies up to 3.675 GHz. The HBv3 also comes with 1920 GiB of local SSD for /scratch.

With up to 480GB per chassis. The HBv3 is ideal for block level workloads in both the front-end and back-end of the design cycle.

HBv3 VMs offer Constrained Cores sizes. This enables customers to pick a lower number of cores to expose to the VM while keeping all other assets constant. Doing so increases how those assets are allocated on a per-core basis. In HPC, common scenarios for which this is useful include:

- Providing more memory bandwidth per CPU core
- Allocating more L3 cache per core for RTL simulation workloads
- Driving higher CPU frequencies to fewer cores in license-bound scenarios
- Giving more memory or local SSD to each core
- Constrain memory to its associated NUMA socket reducing memory latency

Compute best practices

For running CPU bound EDA tools, it is recommended to disable hyperthreading.

A core is a sub-section of a processor that has a complete set of functionalities. Cores are generally fully independent of each other. Cores can be physical (pCore) or virtual (vCore). Azure a pCore is equal to 2 vCores

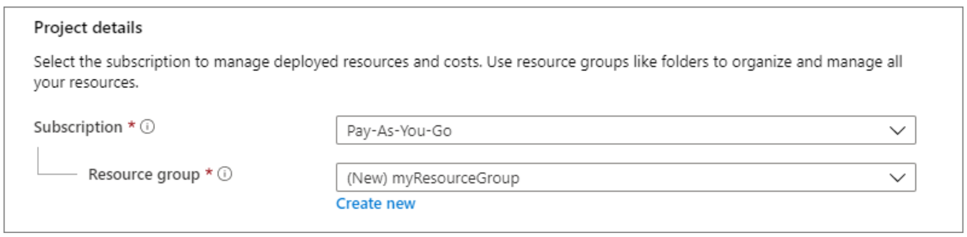
Hyperthreading is a technique for splitting a single physical core (pCPU) into 2 virtual cores (vCPUs). Since much of the core's processing power is shared, disabling hyperthreading is recommended in compute intensive workloads commonly found with many EDA front end tools so that the net processing power is not doubled.

Create a Linux virtual machine in the Azure portal.

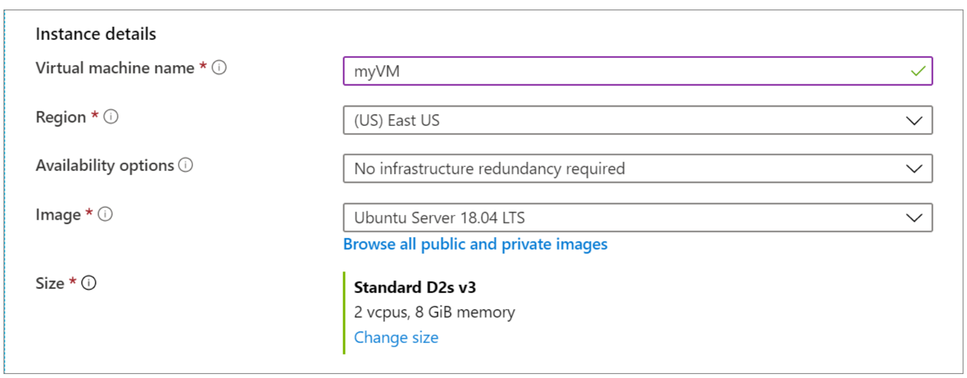
Azure virtual machines (VMs) can be created through the Azure portal. The Azure portal is a browser-based user interface to create Azure resources.

Create virtual machine

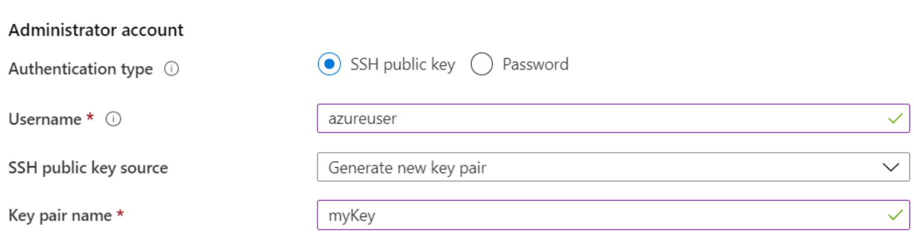
1. Type **virtual machines** in the search.
2. Under **Services**, select **Virtual machines**.
3. In the **Virtual machines** page, select **Add**. The **Create a virtual machine** page opens.
4. In the **Basics** tab, under **Project details**, make sure the correct subscription is selected and then choose to **Create new** resource group. Type *myResourceGroup* for the name.*.



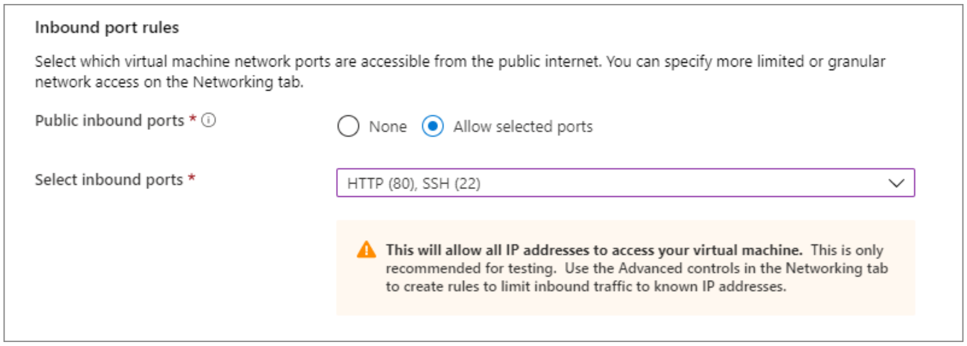
5. Under **Instance details**, type *myVM* for the **Virtual machine name**, choose *East US* for your **Region**, and choose *Ubuntu 18.04 LTS* for your **Image**. Leave the other defaults.



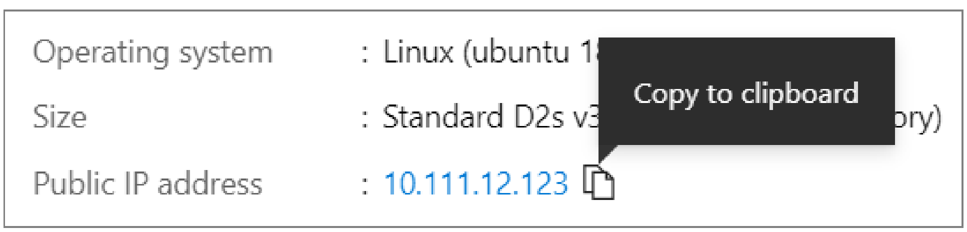
6. Under **Administrator account**, select **SSH public key**.
7. In **Username** type *azureuser*.
8. For **SSH public key source**, leave the default of **Generate new key pair**, and then type *myKey* for the **Key pair name**.



9. Under **Inbound port rules** > **Public inbound ports**, choose **Allow selected ports** and then select **SSH (22)** and **HTTP (80)** from the drop-down.



10. Leave the remaining defaults and then select the **Review + create** button at the bottom of the page.
11. On the **Create a virtual machine** page, you can see the details about the VM you are about to create. When you are ready, select **Create**.
12. When the **Generate new key pair** window opens, select **Download private key and create resource**. Your key file will be download as **myKey.pem**. Make sure you know where the .pem file was downloaded, you will need the path to it in the next step.
13. When the deployment is finished, select **Go to resource**.
14. On the page for your new VM, select the public IP address and copy it to your clipboard.



Shared Image Gallery

A Shared Image Gallery simplifies custom image sharing across your deployment. Custom images are like marketplace images, but you create and customize them yourself. Custom images can be used to bootstrap deployment tasks like preloading tools, configurations, and other OS kernel parameters. Gallery images are versioned, allowing for change tracking and management. A key feature of gallery images is the ability to create multiple copies on high performance azure storage to accelerate deployment at scale.

Storage

EDA tools have their start as desktop applications which wrote to local disk. When these tools moved to a compute grid, we leveraged NFS for network storage because, being POSIX compliant, they functioned like local disk to the tools. Over time, as workloads became more complicated and the scale of EDA workloads grew, even on-premises NFS architectures struggled to keep up. The EDA community experimented with other file systems such as Lustre, GFPS, etc., but, for a variety of reasons, found them unsuitable and continued to use NFS.

However, as EDA workloads continued to increase in complexity, even the best on-premises NFS solutions struggled. This problem only gets worse as customers can exploit the scale of compute available on the cloud. As long as EDA tools continue to rely solely on POSIX file systems, successfully exploiting the near infinite compute available on the cloud for EDA will require that we adapt our methodology to make use of the various options available on the cloud.

To accommodate the demands that extreme scaling makes possible on the cloud, a strategy that combines local attached storage and network storage is required.

Working closely with internal engineering and external ecosystem providers such as NetApp and Pure storage, Azure has complete reference architectures for implementing the best fit of cloud and hybrid storage to provide the optimized NFS infrastructure for EDA workflows. NFS blob allows for cost effective and resilient archiving of historical design data while preserving the instant access the engineers require.

Local Disk

The ability to dynamically choose compute options opens the opportunity to “right size” the compute. This doesn’t only mean the processor and memory, but also storage options. Many Azure VM families have members that include attached NVMe SSD storage. Using this storage option for temporary scratch rather than network disk greatly alleviates stress on the network filer. Especially for workloads like simulation where a large volume of small, random files is generated during the execution phase. This is a strategy that many of the very large fabless customers already employ on their simulation farms.

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 ONTAP performance that organizations run on premises. In addition, deploying the service takes minutes, just select one of three desired service levels and you’re ready to mount a volume with unmatched performance and built-in advanced data management. The 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[10]

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

Microsoft and NetApp have partnered for decades to build increasingly powerful and flexible replication and migration features. For example, Azure 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).

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

The Azure NetApp Files replication functionality provides data protection through cross-region volume replication. You can asynchronously replicate data from an Azure NetApp Files volume (source) in one region to another Azure NetApp Files volume (destination) in another region. This capability enables you to failover your critical application in case of a region-wide outage or disaster.

Azure 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."

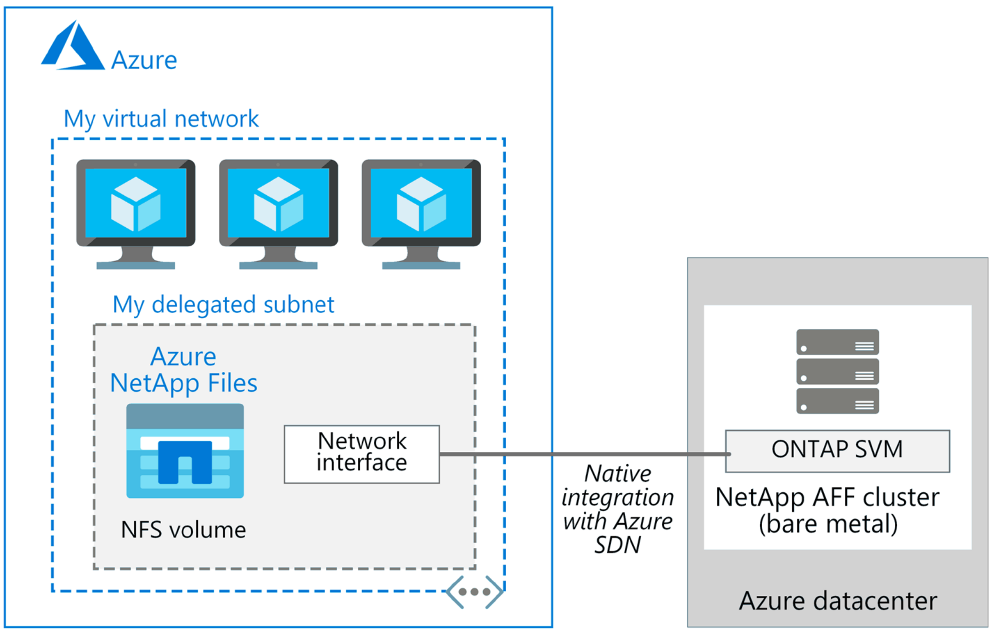


Figure 7. NetApp Files is a fully managed service that is deployed to an Azure subscription.

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

Azure HPC Cache

As we further parallelize EDA workloads, the jobs running on the same design all access the same data at the same time. This can result in thousands of jobs accessing the same file at the same time, causing read contention on the filer.

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 16 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, Blob 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 data center. 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 NFS Blob storage

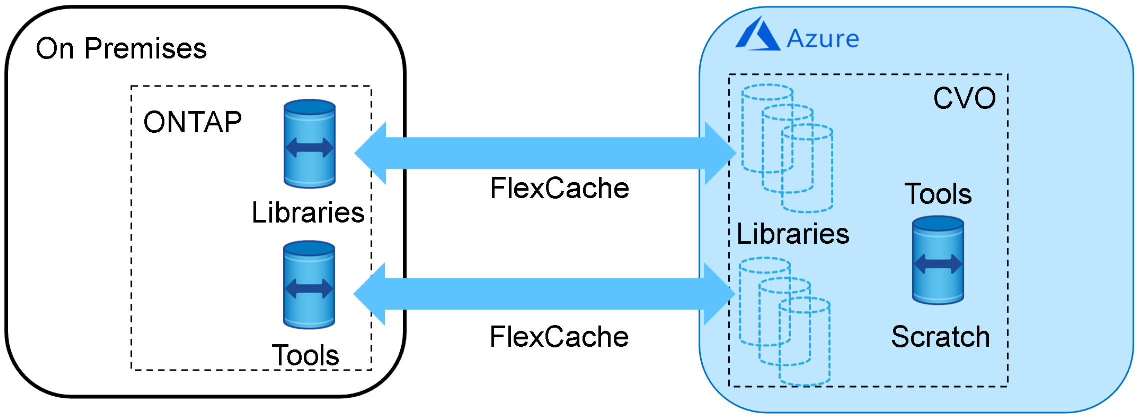
Azure Blob storage is a massively scalable, cost-effective alternative to on-premises solutions for cold data storage. Optimized for unstructured data, Blob storage provides exabytes of capacity with both NFS and REST-based object storage for long-term archives.

Many organizations have years of design data, verification results, etc. that are infrequently accessed, but still need to be easily accessible. Most organizations currently store this data on older, less performant NAS filers. Azure Blob provides a more cost effective and robust, live-archive alternative while still allowing the engineer direct access to the data.

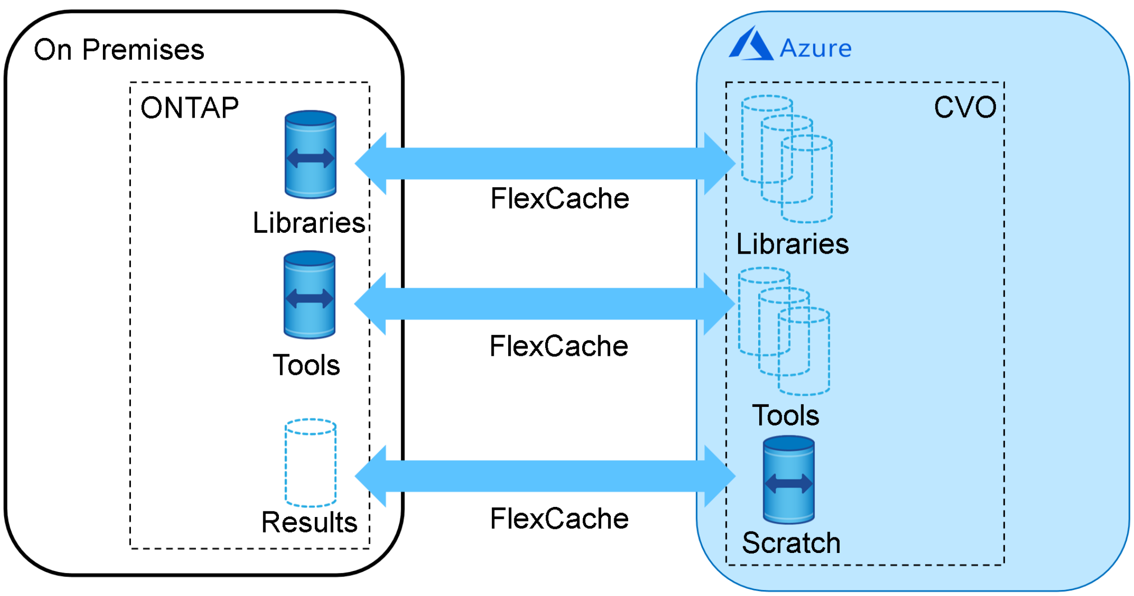
NetApp Cloud Volumes ONTAP

For customers requiring storage in regions without Azure NetApp Files or deploying a hybrid cloud with NetApp systems on premises, Cloud Volumes ONTAP offers ONTAP on Azure.

Cloud Volumes ONTAP can be configured to act as both high-performance storage for EDA workloads in the cloud and a cache for tools and libraries on premises. To an EDA workload in the cloud, tools and libraries appear to be local. Not only is there no need to mirror all the tools and libraries to the cloud, but there is no need to actively manage a separate collection of versioned tools and libraries in the cloud. Get just the data you need, where and when you need it.

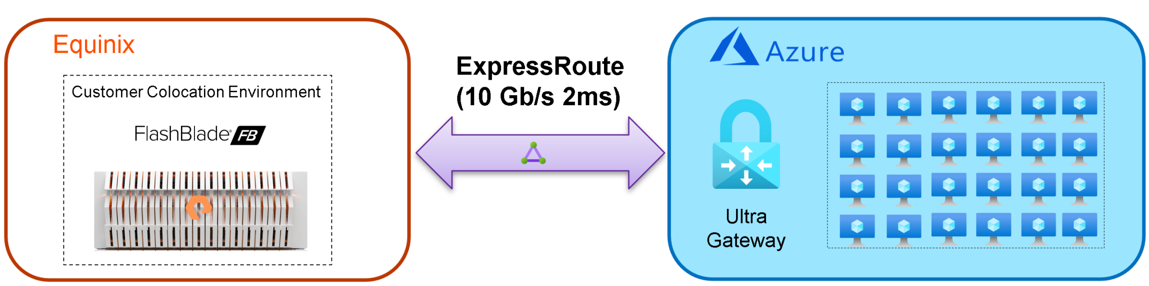


Customers can also configure a "reverse cache." In this configuration, there is an additional cache of the scratch volume on premises. This way, engineers can launch batch jobs in the cloud and debug them on premises as they would if the job was run on premises.



Connected-Cloud with Pure Storage FlashBlade and Equinix

Pure Storage®, Azure and Equinix have collaborated to develop an industry-first solution that allows for customers to run EDA workloads in the cloud while still maintaining full control over security, governance, and sovereignty of their data on-premises or at a colocation facility. The solution is underpinned by Pure Storage FlashBlade as the storage layer connected by ExpressRoute to Azure Compute Instances from a co-located Equinix data center.



Organizations have the option of provisioning Pure Storage FlashBlade through Equinix Metal in a true cloud-like consumption model, or they can acquire the storage unit directly and connect it via Express Route from their own data center or an Equinix colocation facility through Equinix Fabric. There are numerous options available based on budget, performance, and capacity requirements.

Data Security and Control

All data resides on storage devices physically located in data centers as specified by the customer. The customer has complete control over how the data is accessed in the cloud

- Fine grained access controls
- Export rules and policies to specify file system access by IP address and netgroups
- Full data path security with at-rest and in-flight encryption
- Data access using secure Kerberos authentication

Pure Storage FlashBlade Performance Profile

FlashBlade storage performance under NFSv3 scales from a minimum of 300K IOPs (4K read) and 7 GB/s for capacities under 200TB with linear performance scaling to over 3 million IOPs and 50 GB/s as capacities are scaled into the multiple petabytes. To realize the higher bandwidths, the appropriate ExpressRoute configuration is required.

Performance at Scale

The FlashBlade UFFO (Unified Fast File and Object) platform provides numerous benefits for running EDA/semiconductor design tools

- Non-disruptive scalability from 7-150 blades, yielding up to 9PB effective on EDA workloads in a single or multiple namespaces
- Storage performance up to 7.5 million NFS metadata IOPS and 75 GB/s throughput
- Multiprotocol support for NFSv3, NFSv4.1, SMB, and Object
- Native support for replication, quotas, snapshots, backups, and advanced analytics

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 data center 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 Esv5-series. Accelerated networking makes it possible to move much of the Azure software-defined networking stack off the CPUs and onto SmartNICs. These SmartNICS accelerate the network stack using hardware acceleration by leveraging the reconfigurable computing capabilities of FPGAs (field programmable gate arrays). This allows us to devote more compute cycles and system resources to the applications.

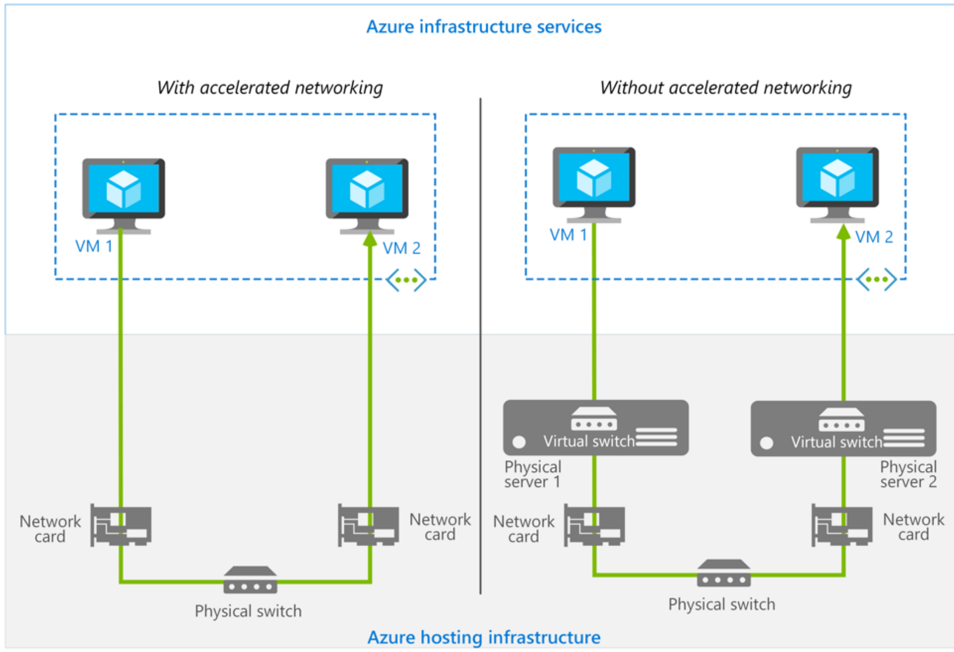


Figure 8. Accelerated networking reduces the number of hops and improves performance.

Networking best practices

To further improve networking performance between compute nodes, proximity placement group (PPG) is recommended

A proximity placement group is a logical grouping used to make sure that Azure compute resources are physically located close to each other. Proximity placement groups are useful for workloads where low latency is a requirement. When the VMs are deployed within the same proximity placement group, they are physically located as close as possible to each other inside of an Azure data center.

Accelerated networking with SRIOV could improve networking performance in many EDA workflows up to 15%, it is recommended to enable it by default.

Azure ExpressRoute

The Azure ExpressRoute network service creates private connections between Azure data centers 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 such as TCAD, schematic capture, etc. are graphically demanding and require low latency between the user and the VM to maintain good user experience. To centralize secure remote access and simplify monitoring and logging, organizations typically deploy a dedicated remote desktop solution. For example, ETX (Exceed TurboX) from OpenText is a popular choice for running EDA with remote desktops on Azure.

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.

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, multiple revisions of design elements, 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 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 job 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

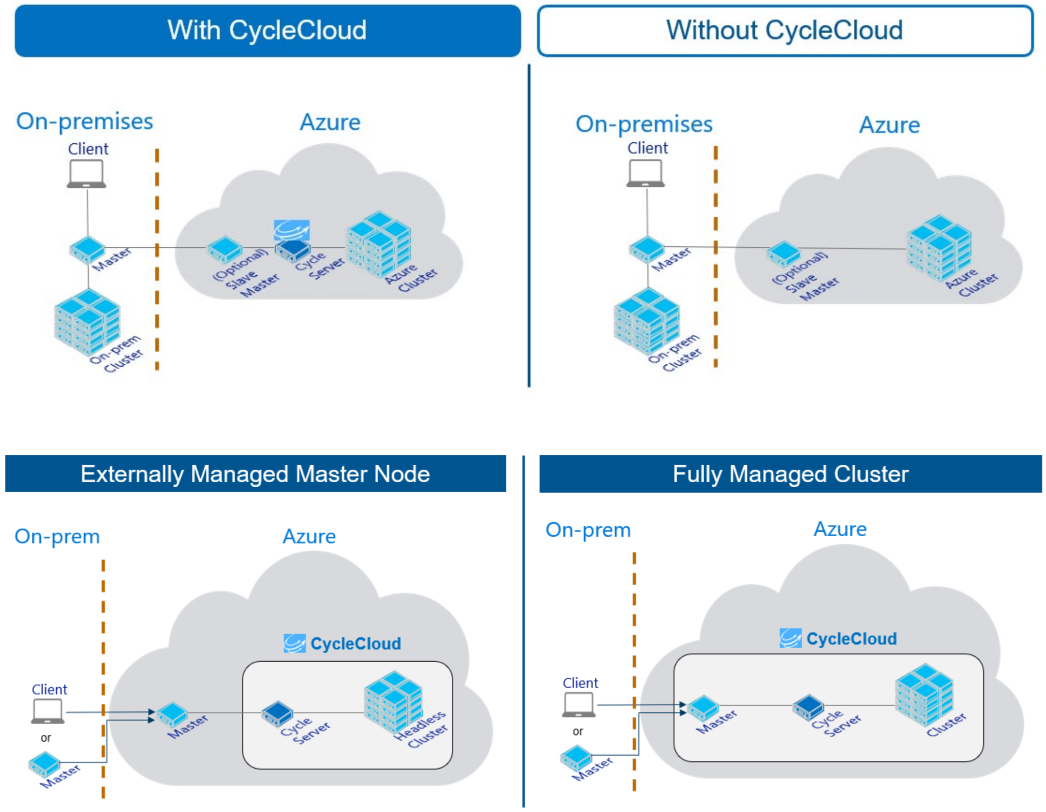
CycleCloud can also orchestrate the deployment of various storage options, including Azure NetApp Files and HPC Cache. Organizations can also define and manage storage components separately or as part of the compute clusters. Administrators and users can manage the data life cycle apart from compute,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

CycleCloud Bursting with LSF

IBM Spectrum LSF is one of the more popular job schedulers for EDA, LSF offers Resource Connector which enables multi-clustering and forwards specific workloads to the cloud based on site defined policies. A typical configuration for the hybrid cloud is to have two LSF clusters, one on premises and another one on Azure, with a master host and execution hosts on each.

Setting up the Resource Connector for bursting workload to the cloud can be a complicated and time-consuming task. CycleCloud can help LSF administrators deploy for a hybrid cloud scenario more easily.

With proven cluster templates and automation scripts, CycleCloud reduces possible errors and the need for human intervention in the creation and operation of a cluster. CycleCloud also supports auto-scaling of the cluster on the cloud, which automatically creates and removes the execution hosts based on actual workloads. CycleCloud's simple web UI allow control and monitoring of the cluster state as well as the Azure costs consumed by the cluster.

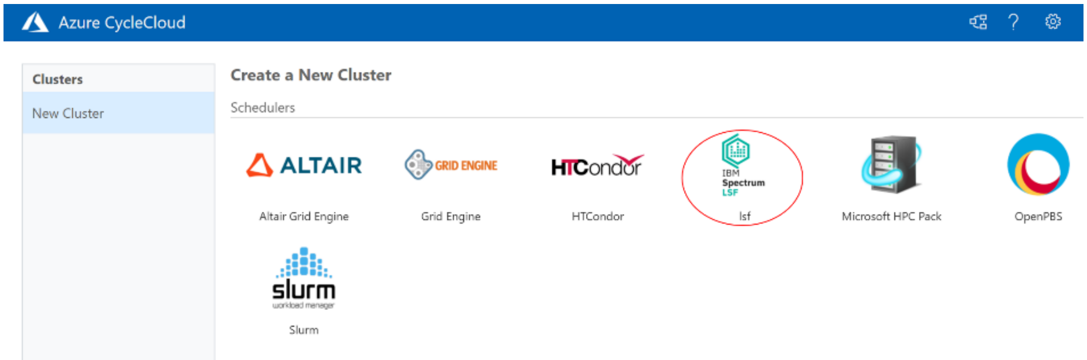


There are two deployment scenarios; 'Externally Managed Master' is recommended for a customized environment, including LSF queues and on-prem NFS. In this case, CycleCloud does not directly manage the LSF controller. A 'Fully Managed Cluster' is managed by CycleCloud directly and can be easily deployed with pre-defined queues and NFS on Azure.

Hybrid LSF with CycleCloud requires IBM Spectrum LSF 10.1 with LSF FP9(532214) or later and CycleCloud version 7.7.4 or later.

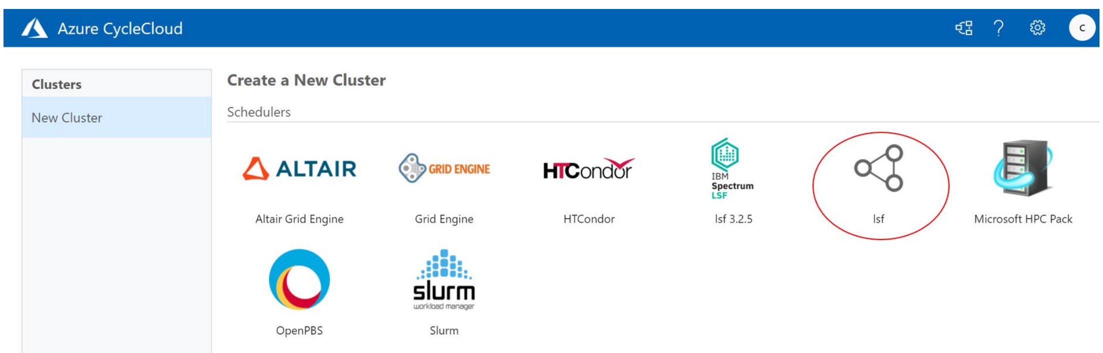
Externally Managed Master

1. Set up a CycleCloud server
2. Create a master VM and configure NFS
3. Install LSF and Resource Connector on the master VM
4. Configure Resource Connector with the following scripts:
    a. cyclecloudprov_config.json: Manages remote administrative functions that the Resource Connector must perform against CycleCloud.
    b. user_data.sh: Defines user-managed proceduers during node startup.
    c. cyclecloudprov_templates.json: Defines the mapping between LSF resource demand requests and CycleCloud instances.
5. Start a headless LSF cluster from CycleCloud Select the LSF template from the CycleCloud UI, enter the required information, and start the cluster. (NOTE: For CycleCloud version lower than 7.9, you can import the headless LSF template from GitHub.)



Fully Managed Master

1. Clone or download the CycleCloud LSF project from GitHub.
2. Copy LSF installers into cyclecloud-lsf/blobs directory.
3. (Optional)Copy a user defined script, for example user_data-full.sh, into cyclecloud-lsf/specs/execute/cluster-init/files directory.
4. Upload the project and cyclecloud-lsf/examples/lsf-full.txt template.
5. Select the uploaded LSF template from CycleCloud UI.
6. Enter required information, including the (user-defined) custom script URI.



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 (CVS).

Part 3: Next steps

Azure offers a systematic approach to support 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 using 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, additional or specialized resources, 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.

We offer cloud workshops and educational services for IT and CAD staff and decision-makers to have a deeper dive into Azure for EDA, the advantages of moving chip design to the cloud, 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.