



The cloud imperative for Semiconductor design workloads



Contents

Introduction	03
Industry context	04
State of EDA on Cloud today	08
Value case for EDA on Cloud	09
Moving to action	12
Conclusion	14

Introduction

Today's market forces have changed the paradigm for semiconductor manufacturers from “innovate to succeed” to “innovate or perish.” Consumers expect the constant rollout of new features and devices.

5G has enabled an intelligent edge, creating demand for specialized silicon for applications like autonomous vehicles and Artificial Intelligence (AI) in the Industrial Internet of Things (IIoT). Use cases such as AI and machine learning (ML) are moving beyond software to hardware solutions, including application-specific integrated circuits (ASICs), field-programmable gate arrays (FPGAs) and purpose-built graphics processor units (GPUs). High tech customers want microprocessors tailored to their specific needs. Meanwhile, transistor counts are soaring to tens of billions per chip. Device design has become impossible without automated semiconductor design software. The challenge is that these applications require massive compute and data storage capacity. As a result, computing models for running automated semiconductor design workloads on premises need to evolve. Chip manufacturers need to explore alternative deployments for their design workloads,

whether that constitutes public cloud, hybrid public/private cloud or a flexible on-premises private cloud.

A public cloud deployment not only brings the benefits of scalability, elasticity and high performance but also a wide array of tools for improving performance, collaboration and project management. It's a good fit for new companies or those who have fully depreciated their private compute assets.

Arm Ltd, for example, is in the process of moving the bulk of its semiconductor design workloads to Amazon Web Services (AWS), for an expected 10-fold increase in semiconductor design and verification throughput.¹

In a hybrid cloud, organizations keep a baseline percentage of compute in-house while moving burst workloads or capacity-intensive steps like register transfer level (RTL) simulations to the public cloud. Moving its burst R&D workloads to AWS enabled NXP to cost-effectively manage demand spikes, growing

its overall compute capacity by up to 50% on an annual basis.² The hybrid-cloud approach allows established companies to leverage their existing infrastructure while taking advantage of the benefits of the public cloud discussed above.

Alternatively, companies can take advantage of flexible on-premises private clouds in which server OEMs assume the cost of installing servers at the facility of the semiconductor client. During high demand periods that would otherwise exceed the company's on-premises capacity, these “cold” servers can be accessed by the client for a fee and turned off after.

The different deployment models all offer varying degrees of scalability and cost benefits. By considering organizational needs and priorities, companies can arrive at the ideal match. The key point is that they move from the traditional model to one that can help them succeed in the modern environment.

Industry context

Table 1: Analog design flow

Phase	EDA Tool Examples	Challenges to Cloud Execution
Schematic/ Circuit Design	Virtuoso Schematic Editor, Synopsys Custom Compiler	Interactive UI, real-time graphics rendering, loading complex design data into tool environment
Front-End Simulation	Cadence Spectre, Synopsys HSPICE	Huge volume of data computation, moving data between on-premises-infrastructure and cloud
Layout Design	Virtuoso Analog Design Environment, Synopsys Custom Compiler	GUI-intensive, detail-oriented physical design data representation with very-low latency requirement
Physical Verification	Cadence Physical Verification System, Synopsys IC Validator, Mentor Graphics Calibre	Real time simulations, complex design rules withery-low latency requirement in data access

Source: Accenture analysis 2021

Electronic design automation (EDA) software is an essential part of chip development, but it is highly resource intensive. The process is complex and involves dozens of applications. Analog design requires different tools and methodologies than digital design (see tables). Different steps, from chip design to packaging to board-level design, each have their own challenges to cloud execution and their own compute needs. To achieve optimal performance, the hardware configuration needs to be tailored to the application.

Table 2: Digital design flow

Phase	EDA Tool Examples	Challenges to Cloud Execution
RTL/Pre-Silicon Verification	Synopsys VCS, Cadence Pegasus Verification System	Huge volume of data computation, moving data between on-premises infrastructure and cloud
Synthesis	Synopsys DC Compiler, Cadence Genus Synthesis Solution	Huge volume of data computation, moving data between on-premises infrastructure and cloud
Formal Verification	Synopsys Formality, Cadence Conformal Equivalence Checker	Huge volume of data computation, moving data between on-premises infrastructure and cloud
Physical Design	Synopsys IC Compiler, Cadence Innovus Implementation System	GUI-intensive, detail-oriented physical design data representation with very-low latency requirement
Physical Verification	Cadence Assura Physical Verification, Synopsys IC Validator, Mentor Graphics Calibre Physical Verification	GUI-intensive, detail-oriented physical design data representation with very-low latency requirement
Static Timing Analysis	Synopsys Prime Time, Cadence Tempus Timing Signoff	GUI-intensive, detail-oriented physical design data representation with very-low latency requirement

Source: Accenture analysis 2021

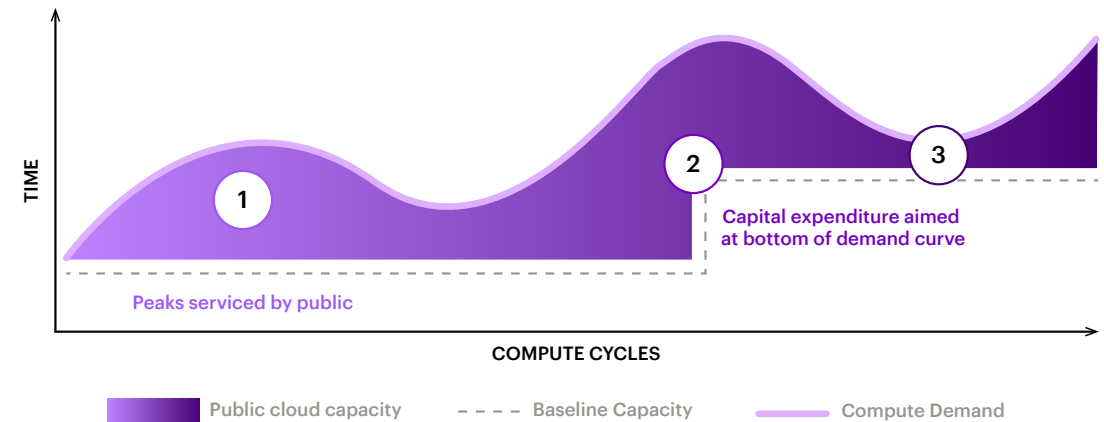
Consider simulation. Two decades ago, a simulation might only have required a handful of CPUs on a single server. Today, bleeding-edge involves partitioning a design to run multiple simulations in parallel across dozens of CPUs and dozens of machines. It’s difficult for any enterprise IT organization and data center to keep up. Meanwhile, if engineering teams don’t get the customized hundreds of resources they need to complete each step in the design flow, development will be slowed down and products will start missing their launch dates. As a result, compute and resource limitations typical have become major bottlenecks in getting products to market.

A semiconductor high-performance compute (HPC) environment includes hundreds of thousands of servers and hundreds of petabytes of storage, expanding by double digits year-over-year. This level of growth takes enormous capital investment, not just for servers but also for the networking equipment, data-center infrastructure, utilities and IT staff to integrate and manage. In addition, users must pay EDA software licensing fees, which scale with infrastructure. Costs mount rapidly and no semiconductor company has infinite resources. The unfortunate result is that design teams find themselves having to make trade-offs between the number of simulations they want to run and the number of simulations they can afford to run.

Restricting semiconductor design workloads to private resources has several other drawbacks. Security in enterprise computing platforms is an ongoing challenge. The value of the intellectual property (IP) involved is incalculable and no organization can afford a ransomware attack. Keeping up with the rapidly evolving threat environment is a never-ending effort that consumes valuable resources.

Finally, even if an organization decides to invest the resources necessary to continually upgrade and maintain a secure HPC infrastructure, there's another nearly unavoidable problem: the ongoing mismatch between resource need and resource availability (see figure 1). Specifying, commissioning, networking and provisioning servers takes time. Meanwhile, engineering teams are competing with one another for capacity and limited in their ability to do their jobs.

Figure 1: Traditional private computing model

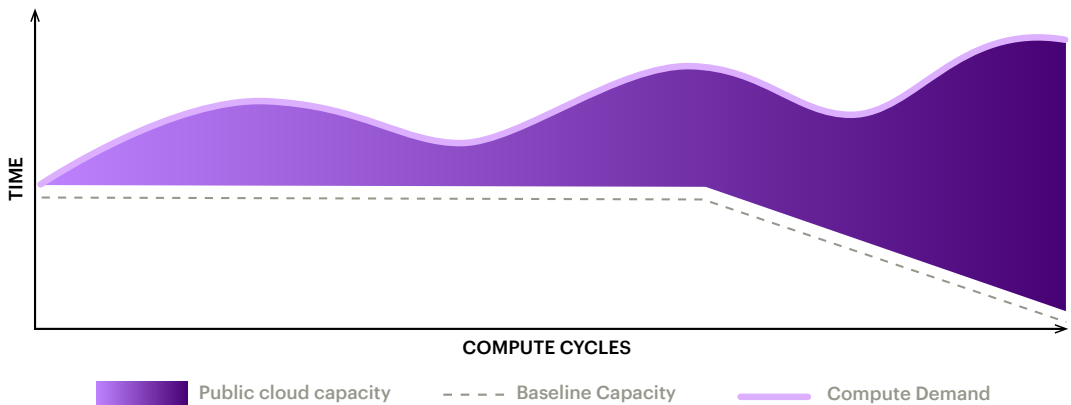


Source: Accenture analysis 2021

In the traditional private computing model, demand and capacity are typically out of sync, with capacity frequently lagging demand.

A hybrid cloud provides a scalable, elastic HPC environment with virtually unlimited compute cycles and storage. The organization can still maintain a certain baseline of compute and storage capabilities. It can use the public cloud for burst workloads or even just the higher-demand phases of the process such as final validation (see figure 2). Engineering teams get immediate access to compute, storage and analytics during needful times. After peak usage, the organization returns capacity to the cloud service provider (CSP) rather than maintaining idle servers until the next surge.

Figure 2: Operating cost approach



Source: Accenture analysis 2021

In the traditional private compute model (figure 1), existing corporate resources provide baseline capacity to address predictable usage while public cloud capacity is instantly available to address burst workloads and cyclical periods of high demand. In the operating cost approach (bottom), the eventual goal is to transition to a fully public cloud or flexible-capacity private cloud deployment.

The public cloud enables users to access fully provisioned servers within seconds, eliminating resource bottlenecks. Performance is part of the business model, which brings big benefits to clients. By using Amazon EC2 z1d instances, for example, which are designed specifically for EDA workloads (they incorporate custom Intel® Xeon® Scalable processors operating at 4.0 GHz), Astera Labs estimates that it cut chip development time by up to 40%.³

Reducing the number of idle servers is also a significant benefit, not just in terms of cost of acquisition but also cost of ownership. With idle servers, license utilization is a concern, since EDA software providers license on a per-core rather than a per-use basis. Idle servers still consume resources, not just power but staff time for updates and patches. Meanwhile, the equipment is aging and losing value and effectiveness with every passing day. A hybrid cloud model enables organizations to take advantage of a variable cost structure that adjusts based on the needs of the moment. They invest their resources in engineering their products, not engineering their data centers.

Some semiconductor companies are committed to maintaining their EDA workloads on premises, maybe because of data egress costs or concerns over security. For those companies, several server OEMs are now offering flexible private compute models that provide cloud-level scalability while reducing total cost of ownership. A company that needs 20,000 servers for its HPC environment, for example, might buy 15,000 of those servers outright. The remaining 5,000 would be installed on premises but owned by the server OEM. They would be networked, provisioned and configured, then turned off. During high demand periods, the company could immediately access these “cold” servers, pay a usage fee while they are operating, then surrender the capacity (and the charges) when the surge is over.

State of EDA on Cloud today

Running semiconductor design workloads in the public cloud is very much the province of early adopters, both in terms of chip manufacturers and the software providers.

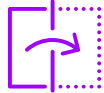
Traditionally, the role of EDA software providers is limited to configuring and deploying their software in client design environment. The problem is that deploying to a cloud environment is very different than deploying to on-premises servers in the corporate data center. Licensing models need to change. Normally, license keys are locked to a specific IP address but cloud workloads can move from instance to instance. EDA vendors have developed interim solutions but still need to work out long-term answers.

In parallel with these workarounds, vendors are also considering what it means to run EDA workloads in the cloud. While EDA applications today run mostly on premises, in the longer term, the greatest value for both EDA vendors and clients would be in the software itself becoming cloud native. Cloud-native software will be able to truly exploit the complete capabilities of the cloud, including the greater cloud ecosystem.

Semiconductor design software vendors are currently in the early stages of this process. They are not yet at the point of producing cloud-native software but in five years, or certainly 10 years, they should be.



Value case for EDA on Cloud



Scalability and flexibility

Probably the biggest single benefit of moving semiconductor design workloads to a hybrid cloud is the near-immediate access to vast amounts of compute and storage resources. Compute demand in chip design varies both cyclically and unpredictably throughout the design process. With automated provisioning and straightforward user interfaces operated by CSPs, a team embarking on static timing analysis or physical verification, for example, can quickly activate as many machines as they need. When the step is over, they return the resources to the cloud and the charges stop.

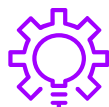
Flexible on-premises private clouds also offer elasticity (although many can't scale down to zero). This model also helps address the latency issue. From a cost standpoint, the OEM bears the brunt of the expense, except when the overflow servers are in active use. On the downside, the semiconductor company is responsible for management and maintenance, plus utilities and floor space (which can be a significant issue in a large-scale computing environment.)



Beyond capacity

The benefits of the public cloud don't stop with scalability. Chip companies are under constant pressure to innovate. To assist them, CSPs offer sophisticated tools and services that are ready for use out of the box. Instead of investing in developing their own AI and machine learning (ML) projects, the engineers can simply use the prebuilt tools and concentrate on their design tasks.

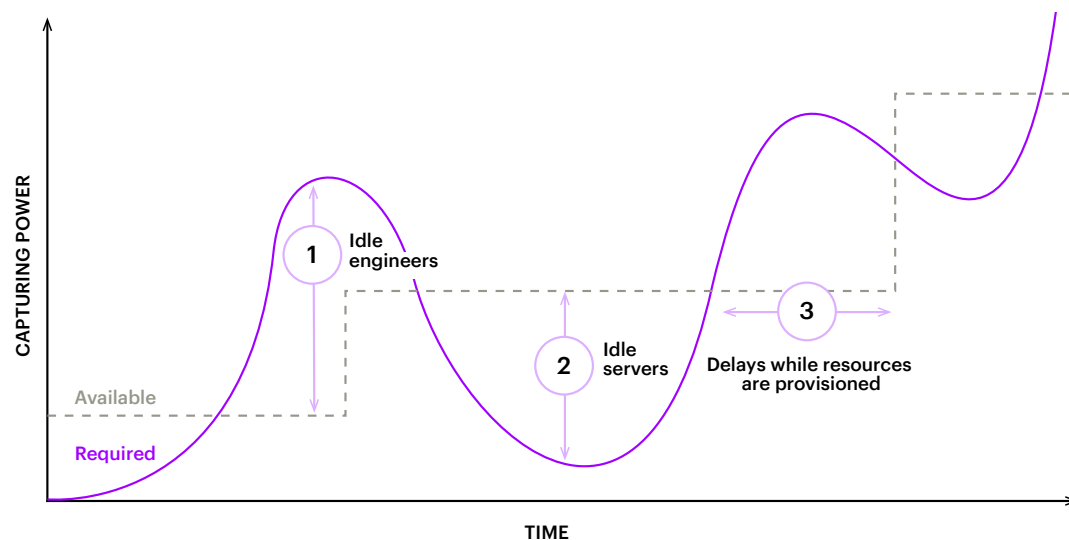
Design teams perform multiple steps simultaneously, using software tools from throughout the EDA ecosystem. Removing storage and compute constraints while adding the power of cloud-native tools, allows organizations to break down the existing silos of R&D, design and build teams and consumers. Cloud-based data hubs help unlock the enterprise-wide data and provide engineers with self-service capabilities to run real-time analytics, as well as more complex AI and ML models. Data curated and prepared once can be used many times across the enterprise to improve innovation.



Productivity

Adding public cloud resources for HPC enables organizations to focus on their core products, not on updating their HPC environments. The chronic mismatch of capacity and demand perpetuated by the traditional model means that for significant portions of the design cycle, either engineers or servers are sitting idle (see figure 3). Engineers sitting idle reduces productivity and delays product launches. Idle servers cost money.

Figure 3: Capacity vs. usage, IT approach



Source: Cadence

Capacity mismatch means that for much of the cycle, either engineers or servers are sitting idle.

Instant access to compute and storage resources makes engineers more productive. By taking advantage of Google cloud AI and ML capabilities, for example, ASML engineers developing an ML tool went from spending hours per day parsing and preprocessing data to almost none. The company estimates that its engineers became 25 times faster at getting data, enabling ASML to accelerate its release timeline from every few months to every two weeks.⁴

Capacity on demand enables simultaneous testing, which can boost efficiency significantly. With the fixed private HPC model, the test and validation process often takes place serially, which can stretch the process out to as long as a quarter. This delay can have wider impact, for example, if a logic block fails a test, that could have ramifications for any other logic blocks that interface with the block under test. With serial testing, delays can cascade out to other sections of the chip. With parallel testing, problems can be addressed more rapidly, minimizing rework.

The combined public and private cloud equips chip manufacturers to reap the advantages of both by allowing data to move between private and public clouds for greater flexibility and continuity in case of platform failure; if there is an outage of the cloud platform, the on-premises platform efficiently runs the workloads.

Predictability and visibility

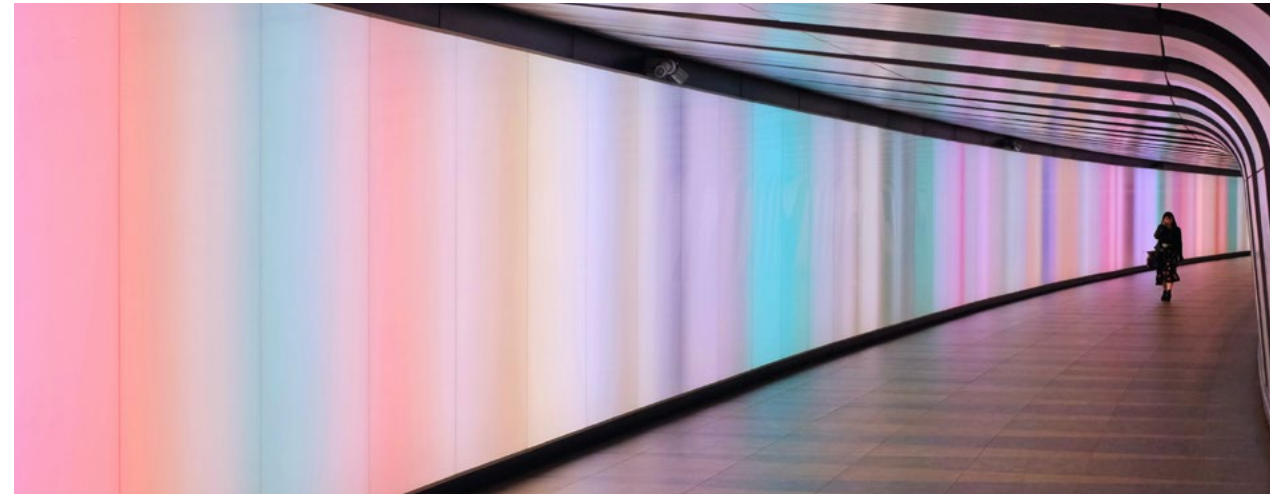
Deploying EDA on a hybrid cloud adds predictability and visibility to the design process. With this approach there is a single integration platform for the entire design flow, a “single pane of glass” to consolidate a formerly disjointed set of tools and methodologies. Individual members can move beyond what is available in an existing report. They can access the raw data and apply analytics and machine learning tools to investigate issues and explore new ideas.

Running EDA workloads on a hybrid cloud presents a more comprehensive view of the various functions, building blocks and in terms of timelines, time for design closure and any potential roadblocks. Comprehensive, detailed input to analytics software improves projections and project management. Greater visibility makes the development process more predictable: Issues are caught early and fixed, products get to market on schedule and the company keeps to its product roadmap. Advance warning of any issues increases the likelihood of successful resolution.

This level of visibility and predictability goes beyond the dashboards and spreadsheets already used by project managers. Once again, the cloud tools come into play, including AI and ML to support intensive process modeling across multiple variables. These capabilities can be used to determine whether a project can be delivered on time or late or what would be required to finish early. The cloud infrastructure provides the tools and visibility to build that level of intelligence into the system.

Security

Security concerns have traditionally been a barrier to public cloud adoption for chipmakers. Traditionally, many thought that data and IP weren’t safe in someone else’s infrastructure. The irony is that the resources expended on security by the average hyperscale CSP far exceeds that of any one enterprise. Given the constantly evolving threat landscape, organizations must play a perpetual game of catch up. It becomes a question of how much a semiconductor company wants to expand dedicated staff and resources versus working with a CSP. Hyperscalers offer a comprehensive suite of automated tools, state-of-the-art best practices like encryption of data in flight and at rest. They have built significant infrastructure in order to protect against threats. For hyperscalers, security is a core competency. Their security teams are often as large as the entire IT staff of the average company.



Moving to action

The choice of deployments for semiconductor design workloads depends on a number of factors, including where the enterprise is in its corporate journey. Established organizations with significant investments in on-premises EDA environments should focus on multistage migrations.

Brand new chip companies (or existing organizations with fully depreciated environments) should seriously consider running EDA workloads purely in the public cloud to take advantage of the many benefits and savings. The key point is that the transition needs to be thoughtful. The migration to cloud, particularly public cloud or on premise flexible private cloud, does not automatically lead to savings and process improvements. Unlocking value requires a strategic approach.

Determine which workloads belong in the cloud and where

Just because a workload can be moved to the cloud doesn't mean it should be. Companies should review each current deployment to understand compute and memory footprints, cost basis and dependencies, as well as identify the workloads that will benefit from the higher level of interoperability, integration and access. As we've discussed, EDA workloads require very specific hardware configurations. They should not be run in a standard instance simply because it is easily available. In the best-case scenario, the instance may be over provisioned for the workload. In the worst case, the workload will fail or underperform because of inadequate resources. It's essential to match the workload to the appropriate instance to get optimal performance.

Be mindful of data cost

EDA workloads can typically involve terabytes of data. Although data ingress is free for most CSPs, there's typically a cost for data egress. Data management should be a key consideration in the decision of which workloads to deploy to the public cloud and which to leave on premises. Depending on the steps and the circumstances of the organization, this could be an argument for a flexible on-premises private cloud.

Focus on value

The choice of architecture should not be considered just a financial and functional decision or just a matter of business strategy, it should be a combination of the two. From a financial perspective, infrastructure is a cost center, a tool for doing business. The enterprise requires a minimum set of capabilities in this area. One argument holds that as long as it's more economical to keep the environment in house and IT can meet SLAs and capabilities, on-premises compute and private data centers can make sense to maintain. As soon as the infrastructure is cheaper when provided by a CSP, companies should begin to move capacity to a public infrastructure.

Focusing on cost alone can be reductive, however. Even before reaching that inflection point, an organization can justify the move from the standpoint of engineering efficiency. Recall the ASML example above. Leveraging the Google cloud tools saved the engineers four hours per day, freeing them up from rote tasks to engage with the creative aspects of the project. The engineers were happier and more effective; the company improved overall efficiency of its development process by an estimated 40%.⁵

Start small

Evaluate the design process and find a key bottleneck that could be addressed using more flexible deployments. Organizations should run a pilot project to learn the process, uncover weaknesses and make any failures on a small scale. Quantitative objectives are essential to know what success looks like.

Don't go it alone

As this paper illustrates, there are many challenges involved in shifting EDA workloads to a cloud environment. Semiconductor companies should work with software vendors, CSPs, system integrators and/or server OEMs for guidance to make this pivot as painless and successful as possible.



Conclusion

For semiconductor companies, infrastructure is just a means for achieving their ultimate goal of designing and building chips. The public cloud is resilient, elastic, secure and mature enough to support even the most sensitive semiconductor design workloads.

Whether leveraged as part of a hybrid cloud infrastructure or a pure-play public option, the public cloud can help companies accelerate design cycles, increase innovation, reduce scheduling risks and keep to product roadmaps. Alternatively, organizations can reap similar benefits of scalability and elasticity from a flexible on-premises private cloud. The most appropriate option varies from company to company but moving EDA workloads to a flexible cloud infrastructure is the logical endpoint for the semiconductor industry. The key point is that chip companies should no longer be focused on building their own on-premises HPC environments for EDA workloads. Infrastructure is not their core competency; they shouldn't consider it a strategic imperative. Their strategic imperative is delivering a steady stream of chips to the market, and that means identifying a migration path to a flexible cloud infrastructure for their EDA workloads.

Accenture has guided hundreds of high tech clients through successful cloud migrations. Reach out to us to get started.



Authors



Syed Alam

Managing Director, Strategy & Consulting,
Semiconductor, Global Lead



Jason Mitchell

Managing Director, High Tech, Cloud Lead, Global



Nick Sidhu

Managing Director, Intelligent Cloud
& Infrastructure, North America



Harmandeep Ahuja

Managing Director, Strategy & Consulting,
Technology Strategy, North America



Harish Ananthamurthy

Associate Director, Technology, Product
Engineering Services, North America



About Accenture

Accenture is a global professional services company with leading capabilities in digital, cloud and security. Combining unmatched experience and specialized skills across more than 40 industries, we offer Strategy and Consulting, Interactive, Technology and Operations services — all powered by the world's largest network of Advanced Technology and Intelligent Operations centers. Our 569,000 people deliver on the promise of technology and human ingenuity every day, serving clients in more than 120 countries. We embrace the power of change to create value and shared success for our clients, people, shareholders, partners and communities. Visit us at **www.accenture.com**.

About Accenture Semiconductor

Accenture Semiconductor is committed to working with semiconductor manufacturers and companies to help capitalize on the opportunities created by digital disruption and optimize efficiencies across product development, manufacturing, supply chain and business operations. We have deep relationships, experience, and expertise across the semiconductor ecosystem: foundries, fabless, equipment makers, and IDM. Along with dedicated practice areas and proven results in mergers and acquisitions, engineering operations, test strategy development, new product introductions, supply chain sourcing/warehouse operations and manufacturing analytics. **www.accenture.com/semiconductors**.

References

1. <https://press.aboutamazon.com/news-releases/news-release-details/aws-and-arm-demonstrate-production-scale-electronic-design>
2. https://d1.awsstatic.com/analyst-reports/AWS+NXP+Case+Study_JM171215.pdf
3. <https://www.intel.in/content/dam/www/public/us/en/documents/case-studies/astera-aws-case-study.pdf>
4. <https://cloud.google.com/customers/asml>
5. <https://cloud.google.com/customers/asml>