THE I/ONX HPC PLATFORM

Justyn Hornor
Chief Executive Officer
I/ONX HPC

January 3, 2026

Abstract

All systems at scale become heterogeneous.

The global compute market is entering a predictable transition. While recent years have been defined by rapid, homogeneous scaling - driven by benchmark competition and short-term performance gains - the next phase will be shaped by optimization across power, cost, supply chains, governance, and lifecycle durability. That phase has not yet fully begun.

I/ONX was founded on the premise that heterogeneity is not a future anomaly, but an inevitable property of scaled systems. As supply chains concentrate, geopolitical risk intensifies, and vendor-specific roadmaps increasingly dictate market outcomes, infrastructure designed around single-vendor dependency becomes a systemic liability. As compute workloads diversify and constraints around energy, supply chains, precision requirements, and geopolitics intensify, optimization must move beyond individual stacks and into infrastructure and policy layers.

The I/ONX HPC thesis asserts that long-term advantage in high-performance computing will accrue to platforms that reduce systemic exposure to supply chain concentration, geopolitical disruption, and vendor lock-in, and that:

- Treat policy and standards as infrastructure
- Enable heterogeneous hardware to evolve without wholesale replacement
- Decouple application intent from underlying hardware implementation
- Align incentives toward long-term system efficiency rather than short-term benchmark optimization

I/ONX does not build vertical end-user applications. Instead, we provide the horizontal integrated platform - hardware, software, and governance - on which the next generation of application developers, enterprises, and sovereign infrastructure builders can deploy and scale heterogeneous systems with confidence.

Executive Summary

From Short-Term Scaling to Long-Term Infrastructure Optimization

The compute industry is approaching the limits of a decade-long phase defined by rapid, homogeneous scaling and benchmark-driven progress [1], [2]. While this approach has delivered extraordinary short-term gains, it has also increased exposure to power constraints, supply chain concentration, geopolitical risk, and long-term architectural fragility [3]–[6].

As workloads diversify and deployment environments become more constrained, optimization can no longer be confined to individual software stacks or accelerator roadmaps. It must move into infrastructure itself - encompassing hardware lifecycles, orchestration and compilation layers, governance frameworks, and economic incentives [7]–[9].

Even if an organization were to commit to a single hardware vendor, the rapid pace of innovation and the increasing diversity of workloads across a given vendor's roadmap would eventually result in a heterogeneous system. The only way to achieve long-term optimization is to treat compute as long-lived infrastructure rather than short-lived products.

The I/ONX HPC thesis began from a simple, first principles, observation: all systems at scale become heterogeneous [8], [10], [11]. Specialization is inevitable. Attempts to enforce long-term homogeneity introduce fragility and lock-in, while unmanaged heterogeneity accumulates operational and governance risk. Durable advantage therefore accrues to platforms that treat compute as long-lived infrastructure rather than short-lived products.

From this premise, the thesis asserts that scalable compute infrastructure must:

- Treat policy and standards as first-class infrastructure components [12]–[14]
- Enable heterogeneous hardware to evolve without wholesale replacement [7], [8]
- Decouple application intent from underlying hardware implementation [11], [15], [16]
- Align incentives toward long-term system efficiency rather than short-term benchmark optimization [9], [17], [18]

I/ONX was built to operationalize this model. By integrating heterogeneous hardware, infrastructure-level orchestration and compilation software, and policy-as-infrastructure through the ASCEND Council, the I/ONX platform enables organizations to deploy and operate compute systems that are adaptable, governable, and durable across changing technologies, markets, and regulatory environments [7], [8], [19], [20].

Rather than competing with the applications built on top of it, I/ONX provides the foundation on which enterprises, scientific institutions, and sovereign infrastructure builders can scale with confidence - reducing systemic risk, preserving optionality, and supporting sustained value creation over decades [9], [17], [21].

I. The Structural Problem: Compute Is Scaling Faster Than It Is Being Optimized

The global compute market has undergone an unprecedented period of growth driven by homogeneous scaling. Performance gains have been achieved primarily through increasing system size, density, and specialization within tightly coupled hardware and software stacks. This approach has delivered short-term results, but it has also introduced structural fragilities that are now becoming visible at scale [1], [3].

At present, the industry remains largely focused on scaling what already exists, rather than optimizing for the conditions that will define the next decade of deployment. As a result, performance benchmarks continue to improve while broader system-level constraints accumulate beneath the surface [1], [4], [22].

A. Homogeneous Scaling and the Limits of Benchmark Optimization

Benchmark optimization has become the dominant signal of progress in high-performance computing and AI infrastructure. While benchmarks provide useful point-in-time comparisons, they increasingly obscure systemic tradeoffs related to power consumption, cost volatility, supply chain exposure, and long-term operability [1], [2].

As systems grow larger and more specialized, the marginal gains from benchmark-driven optimization come at the expense of flexibility. Architectures optimized for a narrow set of workloads become difficult to adapt, expensive to evolve, and fragile in the face of external shocks. What appears efficient in the short term often compounds inefficiency over the full infrastructure lifecycle [1], [22].

This dynamic signals a transition point: the constraints shaping system performance are no longer confined to software stacks or individual accelerators, but extend across physical infrastructure, procurement models, and governance frameworks [12], [13].

B. Supply Chain Concentration and Geopolitical Exposure

Compute infrastructure is increasingly shaped by concentrated supply chains and geographically centralized manufacturing. These dynamics introduce forms of risk that cannot be mitigated through software optimization alone [5], [6].

Single-vendor roadmaps now influence not only performance characteristics, but also precision support, deployment timelines, and long-term availability. At scale, dependence on narrowly sourced components amplifies exposure to geopolitical disruption, export controls, energy availability, and logistics constraints [5], [6].

For enterprises, governments, and infrastructure operators alike, these risks translate into uncertainty around cost, continuity, and control. Systems optimized exclusively for homogeneous scaling lack the structural flexibility required to absorb these disruptions without significant reinvestment [5], [22].

C. Innovation, Regulation, and the False Tradeoff

A common assumption in the market is that regulation and standards necessarily slow innovation. This belief is often reinforced by extractive models that prioritize rapid deployment and short-term returns over durability and alignment [12], [14].

I/ONX takes a different view. At scale, the absence of governance becomes a limiting factor rather than an accelerator. Innovation that cannot be validated, certified, or integrated predictably into existing systems accumulates friction instead of value [12], [13].

To address this gap, I/ONX has developed a broader policy-as-infrastructure framework through its ASCEND Council business line, designed to encode standards, governance, and long-term incentives directly into how compute systems are built and evolved. Within this framework, Certification Labs serve as one operational mechanism for enabling rapid absorption of innovation at the infrastructure level. These labs provide a controlled environment where new hardware, software, and configurations can be evaluated, validated, and integrated without destabilizing operational systems [13].

Rather than slowing progress, certification functions as an acceleration mechanism - reducing downstream risk, shortening deployment timelines, and enabling organizations to adopt new technologies with confidence [12], [13]. When applied consistently, internationally aligned standards also enable cross-border commerce by creating shared technical expectations, reducing friction between jurisdictions, and allowing innovation to move across borders without renegotiating foundational assumptions at each boundary [12], [14].

This approach also creates space for incentive realignment. Organizations that take a longer-term view of infrastructure design benefit from reduced uncertainty, lower lifecycle costs, and faster paths to scaled deployment. In this way, governance and innovation are not opposing forces, but mutually reinforcing components of durable system design [13], [14].

Section 1 establishes the conditions that make a platform-based, heterogeneous, and policy-aware approach not only viable, but necessary. The next section introduces a new model for how compute infrastructure must be designed, governed, and optimized over time.

II. A New Model: Compute as Long-Lived Infrastructure

As compute systems scale, their defining challenges shift. Performance remains important, but it is no longer the sole or even primary constraint [3], [4], [22]. Power availability, supply chain resilience, regulatory alignment, lifecycle cost, and operational durability [7], [23] increasingly determine whether systems can be deployed and sustained at scale.

This reality requires a reframing of compute infrastructure - not as a collection of discrete products or stacks, but as long-lived infrastructure comparable to energy grids, transportation networks, and telecommunications systems [7], [23].

A. All Systems at Scale Become Heterogeneous

Heterogeneity is not an architectural preference; it is an emergent property of scale [10], [11].

As systems grow, they inevitably incorporate multiple processor classes, precision profiles, power envelopes, deployment environments, and operational constraints. Attempts to enforce long-term homogeneity introduce fragility, forcing artificial uniformity onto problems that naturally diversify over time [10].

In practice, large-scale systems already operate heterogeneously - across generations of hardware, mixed workloads, and varied operational contexts. What is often missing is a platform designed explicitly to manage this reality [8], [11].

B. From Product Cycles to Infrastructure Lifecycles

Traditional compute markets are organized around short product cycles and rapid replacement. This model optimizes for near-term benchmarks and feature differentiation, but performs poorly when evaluated across the full lifecycle of infrastructure [1], [7].

Long-lived infrastructure demands a different optimization function that enables component-level upgrade paths:

- Evolution without wholesale replacement
- Predictable upgrade paths across hardware generations
- Compatibility across jurisdictions and regulatory environments
- Capital efficiency measured over decades, not quarters [7], [22]

When compute is treated as infrastructure, success is defined less by peak performance and more by sustained utility, adaptability, and resilience.

C. Policy, Hardware, and Software Must Co-Evolve

In long-lived systems, policy cannot be layered on after deployment. Governance, standards, and incentives shape architectural decisions as directly as hardware capabilities or software abstractions.

Separating technical design from policy considerations creates misalignment: systems that perform well in isolation but fail to integrate across organizations, borders, or regulatory regimes [12], [14], [19]. Conversely, rigid policy disconnected from technical realities constrains innovation.

The alternative is co-evolution. Policy-as-infrastructure provides a stable framework within which hardware and software can innovate without introducing systemic risk [12], [13], [19]. Standards establish shared expectations; orchestration and compilation software provide flexibility; heterogeneous hardware supplies the raw capability to meet diverse workloads [10], [11].

D. Infrastructure Optimization as a Competitive Advantage

As the market transitions from homogeneous scaling to heterogeneous optimization, competitive advantage shifts accordingly.

Benchmark optimization rewards short-term performance gains within narrowly defined conditions. Infrastructure optimization compounds value over time by reducing uncertainty, lowering lifecycle costs, and enabling systems to adapt as constraints evolve.

Organizations that optimize for infrastructure-level efficiency - across power, supply chains, governance, and interoperability - are better positioned to operate at scale in an increasingly constrained and interconnected world [4], [6], [7], [23].

Section 2 establishes the conceptual foundation for the I/ONX platform. The sections that follow describe how this model is operationalized through hardware, software, and policy components designed to function together as a unified system.

III. The I/ONX Platform: Operationalizing Infrastructure Optimization

The preceding sections establish why compute infrastructure must be treated as long-lived, heterogeneous, and policy-aware [7], [8], [19]. Section 3 describes how I/ONX operationalizes this model through an integrated platform composed of four interlocking product lines.

Rather than offering isolated products optimized for narrow use cases, the I/ONX platform is designed as a cohesive system. Hardware, software, and policy components are intentionally co-developed so that advances in one layer reinforce, rather than destabilize, the others [7], [8], [19].

This approach allows I/ONX to address a wide range of customer needs without forcing uniformity. Different organizations engage different parts of the platform based on their priorities, constraints, and regulatory environments. The platform is modular by design, yet unified by a common architectural and governance framework.

At its core, the I/ONX platform is built to support three outcomes:

- Enable heterogeneous systems to evolve over time without wholesale replacement
- Decouple application development from underlying hardware constraints
- Provide predictable, governable pathways for innovation at scale

The following subsections describe each product line and its role within the broader platform.

A. Hardware: A Certifiable Heterogeneous Compute Ecosystem

Hardware decisions within the I/ONX platform are made explicitly in service of infrastructure optimization - prioritizing lifecycle efficiency, evolvability, and systemic resilience over short-term, component-level gains [4], [8], [22]. This prioritization is in service of the application stack and the necessary performance outcomes, such as time-to-first-token and inference latency.

I/ONX hardware is designed from the outset as long-lived integrated infrastructure rather than as a collection of discrete devices. The goal is not to optimize individual components in isolation, but to construct a hardware ecosystem that can evolve predictably as workloads, technologies, and external constraints change - while still delivering the necessary performance outcomes.

This ecosystem-centric approach enables heterogeneous systems to be deployed, operated, and upgraded without forcing rip-and-replace resets or vendor-dependent replacement cycles.

1) Hardware as Infrastructure, Not Devices: In conventional compute markets, hardware is often treated as a consumable - selected to maximize near-term performance and replaced wholesale as new generations emerge. This model performs poorly at scale, where power availability, physical infrastructure, and capital planning impose constraints that extend far beyond individual devices [3], [4], [22].

I/ONX treats hardware as infrastructure [7], [8]. This means designing systems around lifecycle durability, interoperability, and governance from the start [7], [8], [13]. Hardware decisions account not only for computational capability, but also for how systems are powered, cooled, serviced, certified, and evolved over time.

By shifting the optimization target from devices to systems, I/ONX enables customers to plan infrastructure deployments measured in decades rather than product cycles [7], [8], [22].

2) Symphony, Synth, and Canon: Distinct Roles Within a Unified Ecosystem: The I/ONX hardware portfolio - Symphony, Synth, and Canon - is intentionally modular and optional by design. Each product serves a distinct role within the broader infrastructure ecosystem and may be deployed independently or in combination, depending on application requirements.

Symphony is designed as a heterogeneous compute platform capable of integrating multiple processor classes within a governed, certifiable system. It provides the foundational compute fabric on which diverse workloads can be executed. In practical deployments, Symphony can support configurations of up to 64 accelerators (as of early 2026) attached to a single head node in a single data center rack, simplifying orchestration while reducing overall memory duplication, storage overhead, and power consumption through the use of headless PCIe-attached nodes.

Synth extends the ecosystem by enabling flexible system composition and integration across specialized accelerators

and compute configurations. Synth is designed as an computing device at the edge that can operate independently or be coupled with Symphony, enabling heterogeneous compute to be deployed closer to data sources while remaining compatible with centralized orchestration and governance. Synth is not required for all deployments, but can be introduced where additional composability, locality, or customization is needed. Use cases include extreme environments (e.g., field robotics, edge AI), specialized accelerators (e.g., AI inference, signal processing), and custom compute configurations (e.g., specialized hardware for specific workloads).

Canon is a high-speed storage appliance engineered as a first-class component of the compute stack. Rather than serving as generic storage, Canon is designed to be tuned to the specific access patterns, throughput requirements, and data lifecycles of the application stack it supports. Ultra-low latency storage is the primary target, with a focus on scaling up to 100 petabytes of storage per node.

Importantly, these components are not prescriptive bundles. Symphony and Synth can operate independently or together, and Canon can be paired with I/ONX compute platforms for ultra-low latency storage or integrated into existing environments where high-performance, application-aware storage is required.

3) Multi-Processor-Class Support by Design: Heterogeneity is a foundational assumption of the I/ONX hardware approach. Systems are designed to support multiple processor classes - including CPUs, GPUs, ASICs, FPGAs, and emerging accelerators - without privileging a single architecture as the default.

This enables organizations to mix processor types based on workload characteristics such as precision requirements, latency sensitivity, throughput demands, and power efficiency. Scientific workloads requiring IEEE FP32 or FP64 precision can coexist alongside accelerators optimized for inference or signal processing.

By designing for multi-processor-class operation from the outset, I/ONX avoids the architectural rigidity that arises when heterogeneity is introduced as an afterthought. This also enables emerging processor classes to be introduced without requiring rip-and-replace system replacements.

4) Power, Cooling, and Physical Constraints as First-Class Inputs: At scale, power delivery and thermal management are among the most significant constraints on compute infrastructure [3], [4], [22]. I/ONX hardware is designed with explicit awareness of rack-level power budgets, local/regional power standards, and diverse cooling strategies.

Systems are engineered to operate across varying electrical environments, including differences in grid frequency, voltage standards, and AC/DC configurations. Cooling considerations - air, liquid, or hybrid - are treated as design inputs rather than deployment-time accommodations.

This approach allows infrastructure to be adapted to available power and physical environments, rather than forcing facilities to conform to narrowly optimized hardware assumptions.

5) Certification-Ready Hardware Architecture: Hardware within the I/ONX ecosystem is designed to be certifiable, auditable, and reproducible. This capability is essential for deployments operating across regulatory regimes, sovereign environments, and industry-specific compliance requirements.

Certification readiness enables hardware configurations to be validated through ASCEND-aligned standards and Certification Labs without freezing innovation. New components and configurations can be introduced, evaluated, and approved without invalidating existing deployments.

This ensures that governance and adaptability coexist, rather than competing.

6) Hardware Evolution Without Wholesale Replacement: Long-lived infrastructure must evolve incrementally [7], [8]. I/ONX hardware architectures are designed to support mixed-generation operation, phased upgrades, and selective replacement where appropriate [7], [8].

By avoiding forced generational resets, organizations can reduce downtime, preserve capital investments, and integrate new capabilities as constraints and requirements evolve. This approach aligns hardware evolution with real-world operational and economic timelines rather than vendor-driven release cycles.

By treating hardware as a durable foundation rather than a disposable asset, I/ONX enables infrastructure optimization that compounds over time rather than resetting with each product generation.

B. Software (Orchestration): Equalizer (EQ) SDK

The EQ SDK is a software platform that provides the control surfaces and substrates required to deploy, scale, and evolve heterogeneous compute systems predictably. It is designed to be a unified interface for a range of accelerators and compute configurations, providing a consistent experience for users and operators.

As heterogeneous systems scale, orchestration becomes the layer that determines whether diversity in hardware results in efficiency or operational friction [8], [11]. EQ SDK is designed as infrastructure-level orchestration software, providing the control surfaces and substrates required to deploy, scale, and evolve heterogeneous compute systems predictably. While immediate focus is on common frameworks like Kubernetes, the I/ONX team is developing a custom substrate that interfaces with common orchestration toolchains for the granular control needed by a range of accelerators.

1) Orchestration as Infrastructure Control, Not Application Logic: EQ SDK is built primarily for Infrastructure Engineers responsible for operating systems at scale. Rather than embedding orchestration logic inside application frameworks, EQ operates at the infrastructure layer - where decisions about

scheduling, scaling, and resource allocation have long-lived architectural consequences [8], [24].

This separation allows Application Engineers to focus on application behavior and outcomes, while Infrastructure Engineers use EQ to manage how those applications are deployed and scaled across heterogeneous compute environments. Orchestration, in this context, becomes a durable system capability rather than transient middleware.

2) Scaling Containerized Applications Across Heterogeneous Compute: EQ SDK integrates with standard, widely adopted infrastructure tooling used to operate containerized applications. Kubernetes is the immediate focus, with support for Helm charts and deployment manifests that allow heterogeneous compute resources to be expressed and managed using familiar operational patterns [25].

While Kubernetes is the current emphasis, EQ is not limited to a single orchestration framework. Support for additional infrastructure-as-code and orchestration platforms, including Terraform and related tooling, is part of the roadmap. This ensures that EQ can adapt as operational standards evolve across organizations and regions.

Through these integrations, Infrastructure Engineers can scale workloads up or down, assign workloads to appropriate classes of compute, and manage heterogeneous clusters without introducing custom, application-specific orchestration logic.

3) Explicit Mapping of Workload Roles to Hardware Classes: A central capability of EQ SDK is the explicit mapping of workload roles to hardware characteristics [8], [11]. Rather than assuming uniform execution environments, EQ allows different components of an application to be scheduled based on their specific performance, precision, and efficiency requirements.

One illustrative example is a multi-agent or expert-based AI architecture, where different components of a system have materially different compute needs. In such cases, backbone models may be best suited to large, high-memory accelerators, while agents, experts, or auxiliary components can execute efficiently on smaller, specialized accelerators.

This example is not prescriptive, but representative of a broader pattern: decomposing workloads into roles that can be matched to appropriate hardware classes. EQ enables this decomposition to be operationalized without hard-coding hardware assumptions into the application itself.

4) Infrastructure Optimization Through Role-Aware Scheduling: By enabling role-aware scheduling, EQ allows heterogeneous systems to be optimized at the infrastructure level rather than through homogeneous over-provisioning. Expensive, high-capability accelerators can be reserved for workloads that truly require them, while more efficient hardware handles auxiliary or scalable components.

The result is improved utilization of premium compute resources, reduced power consumption, and lower capital expenditure per deployed workload [3], [4], [8], [22]. Importantly, these gains can be achieved while maintaining near state-of-the-art performance, as system-wide efficiency improves even when individual components are optimized differently.

This approach exemplifies the I/ONX shift from benchmark optimization to infrastructure optimization - using orchestration to align workload structure with the realities of operating heterogeneous systems at scale.

5) Incremental Deployment and Evolution: EQ SDK is designed to support incremental adoption rather than requiring wholesale architectural change. Most organizations do not transition to heterogeneous systems in a single step; instead, new compute capabilities are introduced alongside existing infrastructure and expanded over time.

EQ enables this evolution by supporting mixed environments in which different generations of hardware, multiple accelerator classes, and legacy systems can coexist under a unified orchestration model [8]. Infrastructure Engineers can introduce new processor types, scale specific workload components independently, and retire legacy resources gradually without disrupting application behavior.

This incremental approach reduces operational risk, shortens time-to-value, and allows organizations to adapt their infrastructure in response to changing workloads, supply conditions, and regulatory environments - without forcing application rewrites or system-wide resets.

6) Orchestration as a Governance and Policy Surface: At scale, orchestration becomes a point of control where technical execution intersects with governance requirements. EQ SDK exposes orchestration as a policy-aware surface, enabling visibility, auditability, and enforcement without embedding policy logic directly into applications.

Through EQ, organizations can:

- Observe how workloads are scheduled and executed across heterogeneous resources
- Enforce constraints related to hardware eligibility, data locality, or operational boundaries
- Support certification and standards requirements without constraining innovation

This capability allows policy-as-infrastructure frameworks, such as those developed through the ASCEND Council, to be operationalized at runtime. Orchestration becomes the mechanism through which standards are applied consistently, across deployments and jurisdictions, while preserving flexibility and performance.

By treating orchestration as both a technical and governance layer, EQ SDK enables heterogeneous systems to scale in a manner that is not only efficient, but predictable and trustworthy.

C. Software (Compilation): Conductor (Optional by Design)

As heterogeneous compute ecosystems expand, compilation becomes a central challenge - not because computation itself

is new, but because the number of accelerators, architectures, and compiler toolchains continues to fragment. Conductor was designed and built by I/ONX to absorb this complexity on behalf of customers, enabling application teams to remain productive in familiar environments while retaining freedom of hardware choice.

1) Compilation as Translation, Not Rewriting: Conductor is built around a simple premise: Data Scientists and Application Engineers should not be required to abandon established workflows in order to take advantage of heterogeneous compute. PyTorch- and CUDA-based applications remain the primary development environments for many organizations, and Conductor is designed to work from those foundations.

Rather than introducing a new programming model, Conductor translates application intent into forms that can be executed efficiently across diverse hardware. This allows teams to focus on application outcomes while insulating them from the complexity of chip-specific compilation paths.

2) From Application Code to a Hardware-Agnostic DAG: Conductor ingests applications written in familiar frameworks and constructs a Directed Acyclic Graph (DAG) that represents the application's computational structure, data dependencies, and execution flow [15], [16], [26]. This DAG serves as a high-level intermediate representation that sits above vendor- and architecture-specific compiler layers [15], [16], [26].

By operating at this level of abstraction, Conductor separates application logic from hardware implementation. The DAG captures what the application needs to compute, without prematurely committing to how or where those computations are executed.

3) Rules-Based Lowering and Reuse of Tuned Kernels: From the DAG, Conductor applies a rules-based lowering process to map operations into existing compiler toolchains and optimized execution paths. Where highly tuned kernels already exist - such as rocBLAS, cuBLAS, and equivalent vendor libraries - Conductor targets these implementations directly, allowing applications to benefit from years of optimization without reimplementation [16].

This approach ensures that performance gains achieved by hardware vendors and open-source communities are preserved rather than bypassed. Conductor complements these toolchains instead of replacing them, coordinating their use across heterogeneous systems.

In addition to kernel reuse, Conductor is capable of fusing compatible operations within the DAG to produce incremental performance improvements. Fusion is applied selectively, improving efficiency where it is beneficial while maintaining correctness and portability [16], [26].

4) Optional Adoption and Integration with Pre-Containerized Applications: Conductor is an optional component of the I/ONX platform. Organizations may choose to deploy pre-containerized applications or rely on existing binaries where appropriate. In these cases, Conductor can be bypassed entirely.

For organizations with custom applications, evolving models, or a desire to experiment across accelerator classes, Conductor becomes a critical toolchain. Its optional nature allows teams to adopt it where it adds value without imposing it universally across the platform.

5) Obfuscating Accelerator and Compiler Fragmentation: Each accelerator vendor typically maintains its own compiler stack, often with significant differences across product generations [15], [16]. Managing these variations directly requires specialized expertise and ongoing maintenance effort that scales poorly as ecosystems diversify.

Conductor obfuscates this fragmentation by providing a stable compilation interface above vendor-specific compilers. Application teams interact with a consistent abstraction, while I/ONX assumes responsibility for integrating new accelerators, managing compiler evolution, and validating execution paths.

This reduces operational burden and allows organizations to adopt new hardware options without absorbing corresponding complexity.

6) I/ONX as the Adopter of Hardware Evolution: A core responsibility of the I/ONX platform is to track and integrate emerging accelerator technologies on behalf of customers. Through Conductor, I/ONX invests in onboarding new chips, tuning compilation pathways, and ensuring compatibility across heterogeneous environments.

This model is explicitly designed to reduce customer adoption risk. By assuming the burden of accelerator onboarding, compiler evolution, and validation, I/ONX enables organizations to select hardware based on performance, efficiency, availability, or architectural fit. This flexibility is typically achieved without requiring application rewrites or changes to development workflows.

Optionality is preserved, adoption risk is signicantly mitigated, and infrastructure evolution can proceed at a pace aligned with organizational needs rather than vendor timelines.

D. Standards & Policy-as-Infrastructure: ASCEND Council

In long-lived infrastructure, standards and policy are not external constraints - they are core system components [13], [14], [19]. The ASCEND Council represents I/ONX's policy-as-infrastructure business line, designed to encode governance, interoperability, and incentive alignment directly into how compute systems are specified, validated, and deployed.

Rather than treating regulation and standards as post-deployment considerations, ASCEND approaches policy as an architectural input. This allows heterogeneous systems to be designed from the outset to operate predictably across jurisdictions, industries, and regulatory regimes.

1) Policy as an Infrastructure Layer, Not an External Constraint: At scale, governance requirements shape system architecture as directly as hardware capabilities or software abstractions. When policy is treated as an afterthought, it

introduces friction, delays deployment, and increases systemic risk.

ASCEND reframes policy as infrastructure [13], [14], [20]. Standards, governance models, and incentive structures are integrated into platform design so that compliance, interoperability, and auditability emerge naturally from how systems are built and operated. This approach enables innovation to proceed without destabilizing long-lived infrastructure.

2) Data Sovereignty and Data Residency as Design Constraints: Data sovereignty and data residency are among the clearest examples of why policy must be operationalized at the infrastructure level [27], [28]. Data sovereignty defines who has authority over data, while data residency defines where data may be stored, processed, and accessed [27], [28].

As AI systems and high-performance workloads scale, data increasingly becomes a strategic national or enterprise asset. Ambiguity around sovereignty or residency creates deployment risk, limits cross-border collaboration, and can force costly architectural redesigns late in the deployment cycle.

ASCEND treats these requirements as first-class design constraints. By encoding expectations around data control and locality into standards and system architectures, I/ONX enables infrastructure that can be deployed confidently within sovereign, regional, and multinational contexts.

3) From Regulation to Operational Controls: Traditional regulatory approaches often rely on static rules and manual enforcement, creating gaps between policy intent and system behavior [20], [29]. ASCEND shifts this model by translating policy requirements into operational controls that can be validated and enforced through infrastructure [20], [29], [30].

This includes standards that define acceptable data handling practices, residency-aware system configurations, and verifiable execution boundaries. Through integration with orchestration and compilation layers, policy requirements can be enforced consistently without embedding regulatory logic directly into applications.

The result is a system in which governance is measurable, auditable, and adaptable - rather than opaque or brittle [20], [24], [30].

4) Certification Labs as an Operational Mechanism Within the ASCEND Framework: Certification Labs are one operational mechanism within the broader ASCEND framework. They provide controlled environments where new hardware, software, and configurations can be evaluated against established standards before being introduced into production systems [31], [32].

In the context of data sovereignty and residency, Certification Labs enable validation of data locality guarantees, access controls, and compliance with jurisdictional requirements. This allows innovation to be absorbed rapidly while reducing downstream risk.

Importantly, Certification Labs are not a gating function. They accelerate adoption by shortening validation cycles, increasing confidence, and preventing integration failures that would otherwise emerge at scale.

5) Enabling Cross-Border Commerce Through Standards Alignment: Consistent, internationally credible standards are a prerequisite for cross-border digital commerce [2], [6], [32]. When technical expectations vary by jurisdiction, organizations are forced to fragment infrastructure or limit collaboration.

ASCEND enables alignment without uniformity. Shared standards establish common technical expectations while allowing local control over data, infrastructure, and governance. This model supports both efficiency-driven economies seeking predictability and diversification-oriented economies seeking sovereign capability development.

By reducing friction at jurisdictional boundaries, ASCEND allows heterogeneous systems to interoperate across borders without compromising sovereignty or security [14], [19], [32].

6) Bridging Technology, Markets, and Governance: ASCEND functions as the connective tissue between the technical layers of the I/ONX platform and the market and policy environments in which they operate. By linking hardware, orchestration, and compilation to standards and governance frameworks, ASCEND ensures that innovation remains deployable at scale.

This policy-as-infrastructure approach allows heterogeneous systems to scale not only technically, but economically and geopolitically - providing the predictability, trust, and alignment required for long-lived infrastructure in an increasingly constrained world.

IV. Customer Archetypes and Platform Engagement

The I/ONX platform is intentionally designed to support diverse customers without forcing uniform adoption. As heterogeneous computing becomes the norm, different organizations face different constraints, incentives, and success criteria [8], [10], [11]. Section 4 describes how three primary customer archetypes engage the I/ONX platform - each pulling on different components based on their objectives.

This segmentation is not a sales taxonomy [7], [19]. While the archetypes are presented distinctly for clarity, they are not mutually exclusive; many organizations exhibit characteristics of multiple archetypes as their technical, regulatory, and market contexts evolve. It is a systems model that informs platform design, roadmap prioritization, and long-term alignment between technical capability and market need. Not every customer uses every part of the platform, and that selectivity is a feature rather than a limitation.

A. Data Centers and Nation-State Infrastructure Builders

Data center operators and sovereign infrastructure builders approach compute as a generational investment [7], [9], [19].

Their primary concerns center on control, resilience, and long-term viability rather than near-term workload optimization [7], [9], [19].

For this archetype, the highest-value components of the I/ONX platform are:

- ASCEND Council for standards, governance, and cross-border interoperability
- Hardware ecosystem (Symphony, Canon, Synth) as a certifiable, evolvable infrastructure substrate

These organizations prioritize:

- Independence from single-vendor ecosystems
- Predictable integration across jurisdictions and regulatory regimes [12], [14], [19], [33]
- Supply chain resilience and long-term availability [5], [6], [9]
- Capital efficiency measured over decades [7], [9], [22]

Certification Labs play a critical operational role for this segment by enabling local validation and integration of new technologies without destabilizing national or regional infrastructure. Through ASCEND-aligned standards, innovation can be absorbed rapidly while maintaining governance guarantees.

In this context, I/ONX functions not as a vendor, but as an ecosystem enabler - providing the technical and policy foundations required to build infrastructure that can adapt to geopolitical, technological, and economic change.

B. Scientific Computing and High-Performance Computing Organizations

Scientific computing and HPC organizations are driven by correctness, determinism, and computational capability [1], [8], [34]. These environments often operate mission-critical workloads where precision, reproducibility, and performance integrity directly affect outcomes [1], [34].

For this archetype, the most critical components of the I/ONX platform are:

- Hardware configurations optimized for precision-first workloads
- EQ SDK for orchestrating heterogeneous systems without sacrificing control

These organizations value:

- Support for high-precision and mixed-precision computation [1], [8], [34] (e.g., IEEE FP32)
- Orchestration across diverse hardware while preserving scientific integrity
- Avoidance of long-term dependency on a single accelerator roadmap
- Integration with existing HPC workflows and operational practices [8], [34]

While standards and certification are assumed internally within many HPC organizations, the platform benefits of I/ONX - particularly heterogeneous orchestration and hardware optionality - address growing constraints related to cost, supply availability, and architectural rigidity.

I/ONX enables scientific organizations to acknowledge the inevitability of heterogeneous compute while maintaining the rigor and reliability their work demands.

C. Generative AI Application Companies

Generative AI application companies - and organizations deploying adjacent AI workloads [3], [4], [22], [35] such as computer vision, segmentation, recommendation, and predictive analytics - operate under intense performance, cost, and time-to-market pressures [3], [22], [35]. Their focus is on throughput, efficiency, and adaptability as models, architectures, and market expectations evolve rapidly. While large language models (LLMs) and related generative systems represent a particularly fast-expanding subset of this category, the underlying infrastructure demands extend across a broad range of AI applications.

For this archetype, the most relevant components of the I/ONX platform are:

- Hardware configurations optimized for AI workloads
- EQ SDK for scaling and orchestrating heterogeneous inference and training pipelines
- Conductor for experimentation, portability, and future-proofing

These organizations prioritize:

- Cost and energy efficiency at scale [3], [4], [22], [35]
- Flexibility to adapt to new model architectures and precision requirements
- Reduced exposure to vendor lock-in as hardware landscapes shift [5], [6], [10]

While standards and policy frameworks are not always a primary buying driver for this segment, they benefit indirectly from the platform's emphasis on interoperability, predictability, and long-term optionality.

Across all three archetypes, the I/ONX platform provides a consistent foundation [7], [8], [10] while allowing engagement to vary based on need. This flexibility enables I/ONX to serve a broad market without compromising architectural coherence or long-term alignment.

V. Platform Sustainability & Alignment

Long-lived infrastructure requires economic models that reinforce durability rather than undermine it [9], [17]. For platforms designed to operate across decades, sustainability is not defined solely by environmental considerations, but by alignment - between customers, partners, policymakers, and the platform itself [9], [17], [36].

The I/ONX platform is intentionally structured so that value creation scales with long-term system performance, reliability,

and adaptability, rather than short-term deployment volume or benchmark outcomes. This approach reflects a core principle of the I/ONX HPC platform: incentives shape behavior, and durable systems emerge when incentives are aligned with lifecycle efficiency [9], [18].

A. Incentive Alignment Over Extraction

Traditional technology markets often reward rapid adoption followed by frequent replacement [37], [38]. While this model can maximize near-term revenue, it introduces volatility, increases total cost of ownership, and amplifies systemic risk for customers operating at scale [9], [37], [38].

I/ONX takes a different approach. By positioning hardware, software, and policy as interdependent infrastructure layers, the platform aligns its economic success with the long-term success of the systems built on top of it. Customers benefit when infrastructure evolves predictably; I/ONX benefits when those systems remain operational, extensible, and relevant over time.

This alignment encourages behaviors that favor optimization across full infrastructure lifecycles - reducing waste, avoiding premature obsolescence, and enabling incremental innovation without forced replacement [9], [17], [18].

B. Revenue as a Function of Infrastructure Value

I/ONX's revenue model is designed to reflect the role of the platform rather than the outputs of individual applications [17], [39]. Because I/ONX does not build or monetize end-user applications, value capture is tied to the infrastructure layers that enable those applications to scale.

This includes:

- Platform access and integration across hardware and software components
- Long-term support, validation, and lifecycle services
- Standards participation, certification, and governance frameworks through ASCEND

By anchoring revenue to infrastructure value rather than application-level success, I/ONX avoids conflicts of interest with customers while maintaining a sustainable business model aligned with system durability [17], [39].

C. Predictability, Trust, and Capital Efficiency

Aligned incentives produce predictability [9], [18]. Predictable systems reduce risk for operators, investors, and governments alike - lowering the cost of capital and enabling longer planning horizons [9], [21].

Through policy-as-infrastructure, standards alignment, and certification mechanisms, I/ONX reduces uncertainty across deployment environments and jurisdictions. This predictability enables organizations to commit capital with greater confidence, accelerates cross-border collaboration, and supports the formation of stable, long-term compute ecosystems [9], [21], [33].

Platform sustainability, in this context, is not an abstract goal. It is a practical outcome of aligning technical design, economic incentives, and governance structures around the realities of operating heterogeneous systems at scale.

VI. Conclusion: Building What Comes After the Current Cycle

The current phase of the compute market has been defined by rapid, homogeneous scaling and benchmark-driven progress [1], [2]. While this approach has delivered extraordinary short-term gains, it has also revealed its limitations as systems grow larger, more interconnected, and more constrained by external realities [3], [4], [6], [22].

The next phase is already underway. As workloads diversify, supply chains tighten, energy and capital constraints intensify, and geopolitical considerations shape deployment decisions, optimization must move beyond individual stacks and into infrastructure itself [3]–[6], [9].

The I/ONX HPC thesis begins with a simple observation: all systems at scale become heterogeneous [8], [10], [11]. From that premise follows a set of requirements that cannot be addressed through incremental tuning or vendor-specific optimization. Long-term advantage will accrue to platforms that treat compute as long-lived infrastructure - designed to evolve, governed to scale, and optimized across full system lifecycles rather than short-term benchmarks [7]–[9], [17].

I/ONX was built to meet this moment. By integrating heterogeneous hardware, orchestration and compilation software, and policy-as-infrastructure through the ASCEND Council, the I/ONX platform enables organizations to deploy and operate compute systems that are adaptable, governable, and durable [7], [8], [11], [20].

Rather than competing with the applications built on top of it, I/ONX provides the foundation on which those applications can scale across changing technologies, markets, and regulatory environments. This platform-first approach reduces systemic risk, preserves optionality, and aligns incentives toward long-term efficiency [9], [18], [21].

As the industry transitions from benchmark optimization to infrastructure optimization, the question facing organizations is not whether heterogeneity will emerge, but whether it will be managed deliberately or allowed to accumulate unmanaged complexity [1], [8], [11].

I/ONX exists to ensure that the next generation of compute infrastructure is built not just for performance today, but for resilience, interoperability, and sustained value in the decades to come [7], [9], [17], [19].

REFERENCES

[1] J. Dongarra *et al.*, "Hpc benchmarking: Past, present, and future," *IEEE Computer*, vol. 52, no. 11, pp. 42–50, Nov. 2019.

[2] MLCommons, "Mlperf training benchmark: Methodology," MLCommons, San Francisco, CA, USA, Tech. Rep., 2022.

[3] P. Patel, "Ai hardware's big problem: Power," *IEEE Spectrum*, vol. 60, no. 1, pp. 34–39, Jan. 2023.

[4] International Energy Agency, *Electricity 2023: Analysis and Forecast to 2026*. Paris, France: IEA, 2023.

[5] Semiconductor Industry Association, "State of the u.s. semiconductor industry," Washington, DC, USA, Tech. Rep., 2022.

[6] C. Miller, *Chip War: The Fight for the World's Most Critical Technology*. New York, NY, USA: Scribner, 2022.

[7] National Academies of Sciences, Engineering, and Medicine, *Bridging the Gap Between Research and Practice in Cyber-Physical Systems*. Washington, DC, USA: National Academies Press, 2022.

[8] J. Shalf, S. Dosanjh, and J. Morrison, "Exascale computing technology challenges," *Proceedings of the IEEE*, vol. 109, no. 3, pp. 435–447, Mar. 2021.

[9] OECD, *Building Resilient Infrastructure: Governance, Finance, and Implementation*. Paris, France: OECD Publishing, 2021.

[10] J. Hennessy and D. Patterson, *A New Golden Age for Computer Architecture*. New York, NY, USA: ACM Books, 2019.

[11] D. Brooks *et al.*, "Heterogeneous computing: Challenges and opportunities," *IEEE Micro*, vol. 40, no. 2, pp. 8–16, Mar.–Apr. 2020.

[12] World Economic Forum, "Global future council on computing: Governance frameworks for scalable digital infrastructure," Geneva, Switzerland, Tech. Rep., 2021.

[13] *Information Technology - Governance of IT for the Organization*, ISO/IEC JTC 1 Std. ISO/IEC 38 500:2015.

[14] OECD, *Digital Security Risk Management for Economic and Social Prosperity*. Paris, France: OECD Publishing, 2022.

[15] C. Lattner *et al.*, "Mlir: Scaling compiler infrastructure for domain specific computation," *arXiv:2002.11054*, 2020.

[16] T. Chen *et al.*, "Tvm: An automated end-to-end optimizing compiler for deep learning," *arXiv:1802.04799*, 2018.

[17] B. Arthur, *The Nature of Technology: What It Is and How It Evolves*. New York, NY, USA: Free Press, 2009.

[18] O. E. Williamson, "Transaction-cost economics: The governance of contractual relations," *Journal of Law and Economics*, vol. 22, no. 2, pp. 233–261, Oct. 1979.

[19] European Commission, "Shaping europe's digital future: Digital infrastructure and governance," Brussels, Belgium, Tech. Rep., 2021.

[20] S. K. Jutoo and V. Krisshnan, "Policy as code: A paradigm shift in infrastructure security and governance," in *Proc. ICGIS 2025*, Oct. 2025, pp. 182–193.

[21] M. Mazzucato, *The Value of Everything: Making and Taking in the Global Economy*. New York, NY, USA: PublicAffairs, 2018.

[22] McKinsey & Company, "The cost of compute: Infrastructure constraints in ai scaling," McKinsey Global Institute, Tech. Rep., 2023.

[23] International Telecommunication Union, *Trends in Telecommunication Reform: Infrastructure as a Service*. Geneva, Switzerland: ITU, 2020.

[24] Kubernetes Documentation, "Auditing," 2025, [Online]. Available: Kubernetes documentation.

[25] ——, "Schedule gpus," 2024, [Online]. Available: Kubernetes documentation.

[26] T. Leary and T. Wang, "Compiled machine learning: Accelerated linear algebra (xla) for tensorflow," in *Curry On*, 2017.

[27] Google Cloud, "Meet regulatory, compliance, and privacy needs (data residency guidance)," 2025, [Online]. Available: Google Cloud Architecture Framework.

[28] D. Burt and C. Brown, "Data residency, data sovereignty, and compliance in the microsoft cloud," Microsoft, White Paper, 2023.

[29] National Institute of Standards and Technology, "Security and privacy controls for information systems and organizations," Gaithersburg, MD, USA, Tech. Rep. NIST SP 800-53 Rev. 5 (Updated 1), 2020.

[30] Open Policy Agent, "Gatekeeper documentation," 2025, [Online]. Available: Open Policy Agent Gatekeeper docs.

[31] *General requirements for the competence of testing and calibration laboratories*, ISO/IEC Std. ISO/IEC 17 025:2017, 2017.

[32] K. Blind, J. M. de la Potterie, and S. K. S. Mang, "Standards and conformity assessment in global supply chains," *Research Policy*, vol. 52, no. 7, 2023.

[33] World Trade Organization, "Digital trade, standards, and regulatory cooperation," Geneva, Switzerland, Tech. Rep., 2022.

[34] TOP500.org, "Hpc system design, benchmarking, and operational practices," 2024, [Online]. Available: TOP500 documentation.

[35] Stanford Institute for Human-Centered AI, "Ai index report 2024," Stanford, CA, USA, Tech. Rep., 2024.

[36] World Economic Forum, "Measuring stakeholder capitalism: Towards common metrics and consistent reporting of sustainable value creation," Geneva, Switzerland, Tech. Rep., 2020.

[37] J. Parker and M. V. Alstyne, "Platform strategy," *Strategy Science*, vol. 1, no. 1, pp. 1–18, 2016.

[38] S. Kaplan and K. Mikes, "Managing risks: A new framework," *Harvard Business Review*, vol. 90, no. 6, pp. 48–60, 2012.

[39] A. Gawer, "Bridging differing perspectives on technological platforms: Toward an integrative framework," *Research Policy*, vol. 43, no. 7, pp. 1239–1249, 2014.

APPENDIX A
CONTEXTUALIZING INDUSTRY TENSIONS AND DESIGN TRADEOFFS

This appendix addresses a set of recurring tensions present in the technical, economic, and policy literature surrounding high-performance computing and large-scale AI infrastructure. These tensions do not represent contradictions in the I/ONX HPC thesis, but rather reflect differences in scope, time horizon, and optimization objectives across communities. Making these tensions explicit clarifies why I/ONX has adopted an infrastructure-first, heterogeneous, and policy-aware design approach.

The purpose of this appendix is not to resolve all debates, but to contextualize them and explain how the I/ONX platform is designed to remain robust in the presence of these unresolved dynamics.

A. Benchmark Evolution vs. Infrastructure Optimization

A common argument within the benchmarking and performance measurement community is that benchmarks can evolve to capture system-level efficiency, incorporating metrics such as power consumption, utilization, or workload diversity. From this perspective, the limitations of benchmark-driven optimization are seen as temporary shortcomings rather than structural constraints.

The I/ONX HPC thesis does not dispute the value of benchmarks or their continued evolution. Instead, it asserts that benchmarks - by design - remain point-in-time abstractions. Even expanded benchmarks struggle to account for long-lived infrastructure considerations such as supply chain volatility, lifecycle durability, governance alignment, and geopolitical exposure. These factors emerge over years or decades and cannot be fully represented by workload snapshots.

From an infrastructure perspective, benchmarks remain necessary but insufficient. They inform component-level decisions, while infrastructure optimization addresses system behavior over time.

B. Heterogeneity as Design Choice vs. Emergent Property of Scale

Some architectural and vendor literature frames heterogeneity as a strategic design choice - an option that organizations may adopt or avoid based on preference or workload characteristics. In contrast, the I/ONX thesis treats heterogeneity as an inevitable property of systems operating at sufficient scale and over sufficient time.

This difference is largely temporal. In short deployment windows or tightly controlled environments, homogeneity can be maintained. Over longer horizons, however, hardware generations diverge, workloads diversify, power and cooling constraints vary by region, and supply conditions fluctuate. Under these conditions, heterogeneity emerges regardless of intent.

The I/ONX position is that the relevant question is not whether heterogeneity will appear, but whether it will be governed deliberately or allowed to accumulate unmanaged complexity.

C. Platform Economics vs. Infrastructure Economics

Much of the platform economics literature is grounded in consumer software and application ecosystems, emphasizing rapid iteration, winner-take-all dynamics, and short feedback loops. These assumptions do not translate cleanly to compute infrastructure, where capital intensity, regulatory oversight, and operational risk dominate.

Infrastructure economics prioritize durability, predictability, and risk reduction over velocity. Value accrues through sustained availability and alignment rather than rapid replacement or extraction. The I/ONX platform is explicitly designed around this infrastructure economic model, aligning incentives with lifecycle efficiency rather than application-level success.

This distinction explains why I/ONX avoids monetization strategies tied to end-user applications and instead anchors value capture to infrastructure capability and longevity.

D. Regulation as Friction vs. Regulation as Enabler

A persistent narrative in technology markets frames regulation and standards as impediments to innovation. This view often reflects early-stage markets where speed of experimentation outweighs the cost of failure.

At infrastructure scale, the absence of governance becomes a limiting factor. Systems that cannot be certified, audited, or predictably integrated accumulate friction rather than value. Standards and certification mechanisms reduce downstream risk, shorten deployment timelines, and enable cross-organizational and cross-border interoperability.

The I/ONX approach treats policy and standards as enabling infrastructure, not external constraints - allowing innovation to proceed within stable, trusted boundaries.

E. Ambiguity in Data Sovereignty and Data Residency Definitions

The literature and regulatory landscape surrounding data sovereignty and data residency remains fragmented. Definitions vary across jurisdictions, industries, and cloud providers, creating ambiguity in enforcement and system design.

Rather than attempting to resolve these ambiguities at the definitional level, the I/ONX platform operationalizes policy requirements through infrastructure. By encoding enforceable controls, locality constraints, and audit mechanisms directly into system behavior, I/ONX enables compliance even when abstract definitions differ.

In this framing, ambiguity reinforces - not undermines - the need for policy-as-infrastructure.

F. Abstraction Skepticism in High-Performance Computing

Segments of the HPC community remain skeptical of orchestration and compilation abstraction layers, favoring tightly controlled, bespoke optimization to preserve determinism and peak performance.

The I/ONX platform does not eliminate this control. Instead, it relocates abstraction to the infrastructure layer, preserving optionality and allowing organizations to adopt orchestration or compilation tooling incrementally. This approach enables heterogeneous systems to scale without forcing uniform adoption of higher-level abstractions.

Abstraction, in this context, becomes a mechanism for control and governance rather than a loss of fidelity.

G. Efficiency Gains vs. Rising Aggregate AI Costs

Empirical data shows that while per-unit compute efficiency continues to improve, aggregate system costs for AI workloads often rise due to increasing model scale, deployment breadth, and demand.

The I/ONX thesis does not claim that infrastructure optimization will eliminate cost growth. Instead, it emphasizes relative efficiency, optionality, and risk mitigation. Infrastructure-first design enables organizations to absorb growth without proportionally increasing fragility, lock-in, or exposure to external shocks.

In this sense, efficiency is measured not only in throughput per watt, but in the system's ability to evolve sustainably under expanding demand.

H. Summary

These tensions reflect genuine, ongoing debates in computing, economics, and policy. The I/ONX platform is intentionally designed to remain viable in the presence of these unresolved dynamics by prioritizing long-term infrastructure behavior over short-term optimization signals.

By making these tradeoffs explicit, the I/ONX HPC thesis aims to clarify the design principles underlying the platform and to provide readers with a framework for evaluating infrastructure decisions in an increasingly constrained and heterogeneous world.