# Why Velox Chose I/ONX: Building AI Infrastructure That Evolves with AI Itself

## Our Stockholm deployment begins at 2 MW and scales to 10 MW of critical IT—on an architecture designed to deliver far more useful AI output per megawatt, with the freedom to choose the right silicon for every workload.

For Velox, selecting an infrastructure partner for its Stockholm AI data center was never simply a decision about which servers to install. AI accelerators, software stacks and supply conditions change every year or two; the buildings that house them are designed to last decades. The real question was how to extract more useful AI output from every megawatt, keep latency low as the system scales, adapt as accelerator technology changes and improve the economics of AI compute over the life of the site.

That combination made I/ONX High Performance Compute the clear choice as system integrator and platform partner. I/ONX brings a density-first, heterogeneous architecture that changes the infrastructure layout itself. Instead of treating conventional server design as fixed, its Symphony SixtyFour platform can interconnect up to 64 hardware accelerators to a single host node over a high-bandwidth, low-latency fabric—reducing host-side overhead, increasing utilization and letting operators match each workload with the compute best suited to it, without locking the facility into a single accelerator ecosystem.

### Stockholm at a glance:

- **2 MW** of AI compute online in early 2027
- **10 MW** of critical IT later in 2027, within a 15 MW facility
- **Up to 64 accelerators** per host node on I/ONX Symphony SixtyFour
- **~1,800 FuriosaAI RNGD NPUs** in Phase 1, alongside GPUs from leading vendors
- **7,000+ NPUs and ~7,000 GPUs** expected at full build-out
- **Sovereign AI**—a locally operated platform for European customers

## A scaling path measured in output, efficiency—and optionality

The Stockholm program is structured to grow in deliberate stages. The first 2 MW of AI compute capacity is expected to come online in early 2027. A further 8 MW is planned later that year, bringing the project to 10 MW of critical IT within a facility with 15 MW of total electrical capacity.

The significance of that path is not only the added capacity. It is the opportunity to prove an operating model at the first deployment, then repeat it at commercial scale. According to I/ONX's modeling, 10 MW built around its architecture can target the useful AI output associated with as much as 50 MW of conventional infrastructure. That is the larger thesis: scale useful compute through density, bandwidth and utilization, rather than simply multiplying traditional server footprints.

Phase 1 is planned to include approximately 1,800 FuriosaAI RNGD neural processing units, operating alongside GPUs from leading vendors. At full build-out, the platform is expected to support more than 7,000 RNGD accelerators and approximately 7,000 GPUs. This heterogeneous model is a key reason I/ONX stood apart: Velox can combine purpose-built inference silicon, GPUs and future accelerators on one platform, selecting for workload fit, latency, availability and economics.

## Changing the infrastructure layout

Traditional AI clusters often inherit a server-by-server architecture in which host CPUs, network paths, power distribution and cooling are repeated as the cluster grows. That repetition can introduce latency, constrain bandwidth between compute resources, consume power that is not directly producing AI work and make future supplier changes more disruptive.

I/ONX starts from a different premise. By consolidating more accelerators behind fewer host nodes, Symphony SixtyFour changes how the floor plan, power envelope and cooling strategy can be considered. For Velox, whose modular data center approach is built around repeatability, speed and lower cost per megawatt, this was the decisive fit.



© 2026 I/ONX, all rights reserved.

"We are not counting accelerators or megawatts in isolation. We are measuring productive AI output per megawatt, over the full operating life of the facility. At 2 MW, we prove the architecture; at 10 MW, the benefits compound across density, bandwidth, cooling and OPEX."

**— Steven Eliuk, CEO, I/ONX High Performance Compute**

## Why heterogeneous infrastructure lowers technology and supply risk

AI infrastructure is being planned against a fast-moving technology cycle. I/ONX's heterogeneous platform gives Velox and its customers more ways to respond: deploying premium GPUs where they are the best fit, using purpose-built NPUs for efficient inference and evaluating new silicon without rebuilding the entire physical environment.

That flexibility matters to capital planning as much as technical design. It reduces exposure to a single supplier, broadens procurement options and allows future purchasing decisions to be based on workload performance, power draw, cooling requirements and total operating cost. It also supports the sovereign AI objective behind the Sweden project, giving European customers a locally operated platform with a wider choice of compute. Together, those advantages made I/ONX more than a systems integrator for Velox—they made it the architectural partner best aligned with the project's long-term economics and risk profile.

"I/ONX builds the way Velox builds: modular, repeatable and ready to scale. That gives our customers performance without lock-in—and if a different accelerator becomes the right answer, the infrastructure is designed to absorb that choice rather than force a redesign."

**— Andrew Sawyer, Co-Founder & CEO, Velox**

## Building the first phase—and the model beyond it

Velox will develop, manage and operate the Stockholm facility under an Infrastructure-as-a-Service model. I/ONX will serve as system integrator and platform partner, providing the heterogeneous architecture that connects and orchestrates the compute environment, while FuriosaAI will provide RNGD silicon, drivers, software and architectural support for high-throughput, low-latency inference.

Over the coming months, the partners will continue technical validation, software integration and supply planning ahead of the first deployment. The immediate milestone is 2 MW. The near-term destination is 10 MW of critical IT. The larger opportunity is far greater useful output per megawatt—achieved through density, connectivity, cooling efficiency and supplier flexibility designed together from the beginning.

**For Velox, that is why I/ONX was the clear choice: not simply to fill a data center with AI hardware, but to build infrastructure that can deliver more useful compute per megawatt and evolve with AI itself.**

## Interested in AI Capacity in Stockholm?

**Velox is now engaging with customers for Phase 1 capacity. Talk to our team: sonali@allianca.group**

**Learn more about I/ONX Symphony SixtyFour: Connect with sales@i-onx.com and visit www.ionxhpc.com**

**About Velox:** Velox develops modular AI data centers built around repeatability, speed to market and lower cost per megawatt, delivered under an Infrastructure-as-a-Service model.

**About I/ONX High Performance Compute:** I/ONX designs density-first, heterogeneous AI infrastructure; its Symphony SixtyFour platform connects up to 64 accelerators from multiple vendors to a single host node.


© 2026 I/ONX, all rights reserved.