VERTIV.

Case study

Firebird AI Factory

Turning vision, technology and execution into AI infrastructure for accelerated deployment and time to first token.



Background

Firebird is a multiple-phase AI infrastructure project developed in Armenia to provide hyperscale computing capacity for artificial intelligence and high-performance computing workloads.

The first phase, inaugurated in August 2026, marked the transition from strategic ambition to operational AI infrastructure. It demonstrated how national ambition, entrepreneurial initiative, technological expertise, passion and strong commitment can converge to deliver a highly complex project within an exceptionally demanding timeframe.

The project also provides a compelling example of the new requirements shaping AI factories. Success is no longer determined by installed megawatts alone. It increasingly depends on how efficiently electrical power can be converted into productive computing capacity—ultimately measured through metrics such as time to first token and tokens per kilowatt.



Challenge

Building advanced AI infrastructure in a demanding location

The project required significant site and access preparation to support construction at this scale and complexity.

Site accessibility represented one of the first major challenges. Some access roads were unsuitable for the heavy vehicles required to transport large prefabricated modules, cooling equipment and other critical infrastructure.

The construction schedule added another layer of complexity. A significant part of the Program had to be executed during the Armenian winter, when low temperatures, snow and difficult ground conditions can materially affect logistics, civil works and installation activities.

In this context, even a relatively small error or delay could have disrupted the entire delivery sequence and compromised the target completion date.

The primary challenge was therefore not simply the availability of equipment. It was the ability to combine:

- Advanced engineering expertise
- Local knowledge and operational capabilities
- Rigorous logistics planning
- Precise sequencing of deliveries
- Coordination across multiple suppliers and contractors
- Synchronisation between civil, electrical and mechanical works
- Rapid decision-making throughout the project

The thermal management challenge

High-density AI systems generate a high concentration of heat. Removing this heat continuously and predictably is fundamental to the performance and reliability of the computing infrastructure.

The cooling system must respond to rapidly changing loads while maintaining the required thermal conditions for critical equipment. Cooling continuity therefore becomes as important as power continuity.

The Firebird architecture combines liquid and air cooling. While liquid cooling addresses the highest-density computing components, air cooling remains necessary for other heat-generating equipment within the data hall.

The result is an integrated thermal-management architecture designed around three priorities:

- Continuity of operation
- High-density heat removal
- Maximum energy efficiency

Optimising tokens per kilowatt

An AI factory must be evaluated not only by the amount of power it consumes, but by the useful computing output generated from that power.

The relevant question is therefore no longer simply, “How many megawatts are available?” but rather:

How much productive AI compute, and ultimately how many tokens, can be generated from every kilowatt supplied to the facility?

Reducing the energy required by supporting infrastructure makes more power available to the IT load.
This directly improves the productive capacity of the site and its overall economic performance.

Thermal management therefore becomes part of the computing value chain.
It contributes directly to the ability of the AI factory to optimise both time to first token and tokens per kilowatt.

The Vertiv thermal architecture

High-temperature cooling and free-cooling operation

Vertiv™ CoolLoop Trim Cooler units were selected to support the demanding operating conditions of the AI factory.

Their ability to operate with elevated water-supply temperatures above 40°C under appropriate design conditions supports dry free-cooling operation, allowing the system to leverage Armenia's external ambient conditions. When outdoor conditions permit, heat can be rejected without relying heavily on energy-intensive mechanical refrigeration. This operating mode can lower the energy required for cooling compared with conventional architectures.

The dry-cooling approach also avoids the continuous consumption of water associated with evaporative cooling technologies, while supporting the project's efficiency and responsible business objectives.

High-performance air cooling

Even in a liquid-cooled AI data centre, a proportion of the thermal load must still be managed through air cooling.

Vertiv™ fan-wall units were therefore deployed to provide powerful and precisely controlled airflow for the components and areas that continue to require air-based heat removal.

These units combine high cooling capacity with advanced electronic controls. They can operate as part of a coordinated local network, using control algorithms designed to balance airflow, cooling capacity, energy efficiency and redundancy across the data hall.

Rather than functioning as isolated pieces of equipment, the units operate as an integrated thermal system capable of adapting to changing AI workloads.



Vertiv™ CoolLoop Trim Cooler

Monitoring and system-level optimisation

A higher-level monitoring and control platform was added to collect operational data and provide coordinated visibility across the infrastructure.

The system supports:

- real-time monitoring of operating conditions;
- early identification of potential anomalies;
- coordinated management of cooling assets;
- optimization of system performance;
- and improved operational continuity.

This system-level intelligence is essential in an AI factory, where rapid changes in computing demand can produce equally rapid variations in power consumption and heat generation.

Delivering the thermal architecture: from equipment selection to coordinated delivery

A decisive success factor was the contribution of ROOT, Vertiv's local representative and distributor in Armenia, whose team operated as an effective extension of Vertiv's engineering and manufacturing organisation.

Application engineers were involved from the initial design phase to identify the most effective and efficient technical architecture for the facility. Their role then extended into delivery planning, installation sequencing and coordination with the different project partners.

This integrated approach connected design, manufacturing, logistics and on-site execution within a single operational framework.

The delivery schedule had to remain closely aligned with the progress of the civil works and with the installation of the mechanical and electrical infrastructure. This required a high level of transparency and continuous communication among all stakeholders.

For a project of this nature, commitment alone is not sufficient. Every participant must bring deep technical competence, because a single misalignment can affect multiple downstream activities.

> "Projects of this intensity are won in execution - a delivery organised and arriving in the right sequence, a decision taken in hours instead of days. Our job was to guide the selection of the right units, to show Firebird the advantages of Vertiv for this project, and of course to make Vertiv's technology land on schedule - facing enormous challenges without a single compromise on quality."
>
> **— Gegham Sargisyan,**
> *General Director, ROOT*

Key success factors

The successful delivery of Firebird's first phase was enabled by the combination of several interdependent factors:

1 A shared strategic vision

Public-sector ambition and private entrepreneurial initiative created the conditions for a project capable of strengthening Armenia's position in the global AI ecosystem.

2 Integrated engineering

The technical solution was developed by considering computing, power, cooling, controls and physical infrastructure as parts of one interconnected system.

3 Strong local capabilities

The local team connected global engineering and manufacturing expertise with the practical realities of the site, the climate and the construction program.

4 Precise logistics and sequencing

Equipment production and delivery were synchronized with civil, electrical and mechanical works, despite limited site accessibility and challenging winter conditions.

5 Advanced thermal technology

High-temperature water operation, dry free cooling, liquid cooling and intelligent air cooling helped minimise supporting energy consumption while maintaining thermal continuity.

6 Commitment supported by competence

The accelerated Program required exceptional dedication, but also the technical expertise to make fast decisions without compromising quality, reliability or safety.

Outcome

The inauguration of Firebird's first phase in August 2026 represented more than the completion of a data centre.

It demonstrated how an AI factory can be delivered in a challenging environment by combining:

- National ambition
- Entrepreneurial vision
- International technology
- Local engineering competence
- Industrial execution
- Shared commitment to a clearly defined objective

The project establishes a scalable foundation for Firebird's subsequent development phases and for Armenia's broader ambition to become a regional centre for advanced AI computing.

Most importantly, Firebird illustrates a fundamental shift in digital infrastructure: the objective is no longer simply to build capacity, but to transform every available kilowatt into the greatest possible amount of useful intelligence.

From megawatts to intelligence

Firebird shows that the next generation of AI infrastructure will be measured not only by scale, but by the effectiveness with which technology, energy and execution are brought together.

The defining metrics of an AI factory will increasingly be: *how quickly it can deliver the first token, how many tokens it can generate per kilowatt, and how it can sustain that performance over time.*

That is where advanced infrastructure becomes a genuine competitive advantage.



© 2026 Vertiv Group Corp. All rights reserved. Vertiv™ and the Vertiv logo are trademarks or registered trademarks of Vertiv Group Corp. All other names and logos referred to are trade names, trademarks, or registered trademarks of their respective owners. While every precaution has been taken to aim for accuracy and completeness here, Vertiv Group Corp. assumes no responsibility and disclaims all liability for damages resulting from use of this information or for any errors or omissions. Specifications, rebates, and other promotional offers are subject to change at Vertiv's sole discretion upon notice.