European High Performance Computing Joint Undertaking

European High Performance Computing

Joint Undertaking

Call for tenders EUROHPC/2026/OP/0009

ACQUISITION, DELIVERY, INSTALLATION AND HARDWARE AND SOFTWARE MAINTENANCE OF INNOVATE INDUSTRIAL GRADE SUPERCOMPUTER – FOR THE EUROPEAN HIGH PERFORMANCE COMPUTING JOINT UNDERTAKING (EUROHPC JU)

Open procedure

TENDER SPECIFICATIONS

Part 2: Technical specifications

Table of Contents

1 Project goals

1.1 Introduction

The INNOVATE project aims to provide Italian and European private organizations with a leading industrial-grade high-performance computing (HPC) system and related services. The primary objective of the INNOVATE project is to design, manage, and operate a flexible HPC system that can provide both computing capabilities and cloud services at Petascale-class, effective for a wide range of industrial applications, including numerical simulation, data analytics, data processing, machine learning/artificial intelligence (ML/AI), and innovation with a vision towards the digitalization of processes and services.

The proposing consortium consists of large industries from various Italian industrial sectors, including Almawave S.p.a, Autostrade per l'Italia S.p.a, Bi-rex S.p.a, CENTRO Nazionale Di Ricerca In High Performance Computing, Big Data E Quantum Computing (Fondazione ICSC), International Foundation Big Data And Artificial Intelligence For Human Development (Fondazione IFAB), SNAM S.p.a, and Unipol Assicurazioni S.p.a. The industrial-grade supercomputer will be hosted in Data Hall number 2 of the data center built and managed by CINECA at the DAMA Technopole in Bologna. This data center was selected through an international competitive procedure to host one of the pre-exascale systems of EuroHPC JU, meeting all requirements in terms of energy efficiency, connectivity, and availability of a high-performance local IT ecosystem. CINECA has a proven track record in designing, managing, and operating high-level HPC infrastructures. The LEONARDO system, hosted at the Bologna Technopole, was ranked 4th in the Top500 list of the most powerful supercomputers in the world. This expertise ensures that CINECA can effectively manage the new industrial grade HPC system, providing high-quality support and services to the consortium partners.

1.2 Goal of the procurement

This project aims to procure a hyper convergent Cloud – HPC based infrastructure able to implement the following main characteristics:

- Provide a Tier-1 class computing HPC system, providing both bare-metal and virtualized services, in order to address the variety of industrial use-cases, focusing particularly on HPC and AI workloads.
- Provide state-of-the-art parallel & multiprotocol Cloud storages, able to serve the compute partitions. It is paramount that workloads will run with adequate quality-of-service to comply with the service level agreements imposed by this critical national cybersecurity service.
- Provide an interconnection able to glue together the Cloud and Orchestrated computing partition and storage partitions, with the flexibility to host Cloud service and HPC driven workloads.

The resources of the DAMA data centre, which were acquired through this procurement, will help address the challenges in industrial digital transformation. In particular, this supercomputing infrastructure will play an essential role in addressing:

- Digital twins.
- LLM model tuning and execution of inference workloads
- Big data analytics for market trends
- Real-time financial transaction processing
- Complex supply chain optimizations
- Advanced robotics and automation
- Personalized healthcare solutions
- High-fidelity virtual prototyping
- Climate modelling and weather forecasting
- High-resolution 3D simulations

Moreover, this procurement targets the provisioning of border apparatus, firewall and network distribution switches to enable a capable external connectivity towards the systems hosted in the DAMA data centre.

1.3 Technical Overview

The system has been designed with flexibility as a core principle, both in terms of management and computational resources. Architecture enables dynamic allocation and orchestration of resources, supporting a variety of operational modes to address diverse and evolving user needs. Multiple CPU and GPU partitions are provided, each optimized for specific workloads such as AI training, inference, visualization, large-scale simulation, and advanced services. This approach ensures that the infrastructure can efficiently support a broad spectrum of applications, from high-performance computing to interactive and service-oriented tasks, adapting seamlessly to future technological developments and user requirements.

Computational resources are divided into two pools, each managed differently, exposing resources to users in distinct ways:

- *Industrial Cloud*: it provides compute resources managed by a cloud computing platform, in these pools will be assigned the 65% of the resources for each compute partition.
- *HPC-O1*: it provides the compute resources managed by an orchestrator platform, in these pools will be assigned the 35% of the resources for each compute partition.

The network configuration must also allow for the computation servers to be moved between modes at any time, ensuring maximum operational flexibility and resource optimization. Each of the two modes will be allocated a certain number of nodes from the following partitions, based on the previously defined percentages:

- *Cloud application*: This partition is designed to host CPUs optimized for service virtualization, ensuring flexibility and scalability for cloud applications. The goal is to provide efficient resources for virtualized environments, micro services, and service-oriented workloads.
- *Industrial HPC simulation*: This partition is dedicated to CPUs tailored for intensive computational workloads, typical of industrial and scientific simulations. It guarantees high performance for numerical calculations, modelling, and large-scale analysis.
- *Scale-out AI/HPC*: the partition is designed to incorporate high-performance GPUs specifically tailored for computational tasks and training AI models. This partition ensures that resources are optimized for intensive AI/HPC workloads, providing the necessary power and flexibility to handle complex algorithms and large-scale data processing.
- *Cloud data driven and Visualization*: This partition is designed to incorporate GPUs specifically oriented towards visualization and inference. The partition ensures that resources are optimized for inference computing needs and real time visualization.

2 Document definitions

2.1 Tendering procedure definitions

Term	Description
Procurer	The entity launching and running the procurement procedure.
Candidate	The qualified economic operator eligible to contract.
Supplier	The tenderer who is awarded the contract as part of this procurement.
Offer	The final bid submitted by the Candidate.

Table 1: Procurement procedure definitions

2.2 Categories of requirements

The requirements and features within the documents are categorized as follows.

Requirements priority	Requirements category	Description
Mandatory	**MRQ**	*Mandatory Requirements.* These specifications constitute essential requirements for the procured infrastructure and must be fully met by all Best and Final Offers. Compliance with Mandatory Requirements will be rigorously evaluated for each submitted Offer. Offers may include enhancements beyond the Mandatory Requirements, such as increased quantity, capacity, or capabilities, which will be positively considered during the evaluation process. Final Tenders failing to satisfy all Mandatory Requirements shall be rejected.
Targeted	**TRQ**	*Highly targeted requirements.* These specifications represent highly desirable features for the procured system. Unlike Mandatory Requirements, failure to meet these targeted specifications will not result in the rejection of the Best and Final Offer submitted by the Candidate.
Mandatory	**DCS**	*Data Centre Specifications.* The Offer shall comply with the detailed data centre specifications. However, within the framework of these requirements, the Candidate is permitted to propose alternative solutions at its own expense, provided that such alternatives

		ensure proper integration of the offered system into the CINECA infrastructure.
Mandatory	**DOC**	*Documentation* All documentation items listed are mandatory and must be provided by all Candidates within their Proposal. Compliance with documentation requirements will be evaluated for each submitted Proposal based on the quality and completeness of the responses provided.

Table 2: Categories of requirements

2.3 Unit of measure

Only SI (International System of Units) decimal prefixes shall be used throughout the Proposal for expressing unit of measure.

Memory & storage capacity:

- 1 kB = 1,000 bytes
- 1 MB = 1,000 kB
- 1 GB = 1,000 MB
- 1 TB = 1,000 GB
- 1 PB = 1,000 TB

Networking Bandwidth (bits per second):

- 1 kbit/s = 1,000 bits per second
- 1 Mbit/s = 1,000 kbit/s
- 1 Gbit/s = 1,000 Mbit/s
- 1 Tbit/s = 1,000 Gbit/s

Storage Bandwidth (data movement rates in bytes per seconds):

- 1 kB/s = 1,000 bytes per second
- 1 MB/s = 1,000 kB/s
- 1 GB/s = 1,000 MB/s
- 1 TB/s = 1,000 GB/s

The compute performance of the system shall be expressed using SI-based floating-point operation metrics:

- 1 kflop/s (KF) = 1,000 floating point operations per second (flop/s)
- 1 Mflop/s (MF) = 1,000 kflop/s
- 1 Gflop/s (GF) = 1,000 Mflop/s
- 1 Tflop/s /TF) = 1,000 Gflop/s
- 1 Pflop/s (PF) = 1,000 Tflop/s
- 1 Eflop/s (EF) = 1,000 Pflop/s

3 Site context

3.1 CINECA hosting entity

CINECA – founded in 1969 – is a not-for-profit Consortium, made up of 117 members: the Italian Ministry of Education, the Italian Ministry of Universities and Research, 70 Italian universities and 45 Italian National Institutions. It is the largest Italian computing centre and one of the most important worldwide. With more than nine hundred employees, it operates in the high-performance computing (HPC), technology transfer and information technology (IT) sectors. CINECA develops advanced IT applications and services with the main goal of supporting academia, public administration, and private companies.

At national and European level CINECA is expected to play a role as advanced research infrastructure provider, bringing its standout experience and expertise in HPC and the ability to support the actions of the centre. In fact, CINECA offers state-of-the-art hardware resources and highly qualified personnel, and is committed to accelerate scientific discovery by continuously evolving its computing, data management and data analysis infrastructure and services. CINECA's HPC infrastructure and expertise support research across all domains, helping in tackle scientific and societal challenges in weather and climate forecasts, computational fluid dynamics, computational bioinformatics, genomics and so on.

CINECA has a proven track record of providing HPC systems at the top of the most powerful computing systems in the world - and three times in the top 10 - as ranked by the top500.org list. CINECA hosts and manages Leonardo, the fourth supercomputing system in the current top500 ranking. The HPC department works for the management, support, and exploitation of the HPC infrastructure, providing services to address computing research needs.

3.2 Project implementation

3.2.1 Procedure

For this procurement, the procurer – and the involved partners – elected to use a public open procedure. The goal of this document is to provide the requirements the supplier is requested to satisfy.

3.2.2 Time schedule

The procurer targets to start the production phase of the procured in 2H of 2027.

3.3 Site visit

During the tendering period, a half-day site visit is planned, tentatively during the last week of August 2026. The exact date, time and practical arrangements will be published in the procurement documents. The purpose of the visit is to provide additional context and support a clear understanding of the data centre environment and the integration requirements for the proposed solution.

Tenderers interested in attending are invited to contact the Contracting Authority at the contact address indicated in the Administrative Specifications to arrange a suitable date and time.

During the site visit, questions will not be answered directly. Any relevant clarifications will be consolidated by the Hosting Entity into an anonymised Q&A document, which will be provided to the EuroHPC JU and published together with the procurement documents on the Funding & Tenders Portal.

4 Data centre facility

The new equipment will be hosted in CINECA data centre located in DAMA Tecnopolo: Via Stalingrado, 96, 40128 Bologna (BO), Italy. To provide guidance on the logistic aspects and data centre integration, this Chapter outlines the data centre specifications and the main architectural design elements.

4.1 Facility description

Figure 1 shows what is currently deployed in the Big Data Technopole (Tecnopolo) data centre.

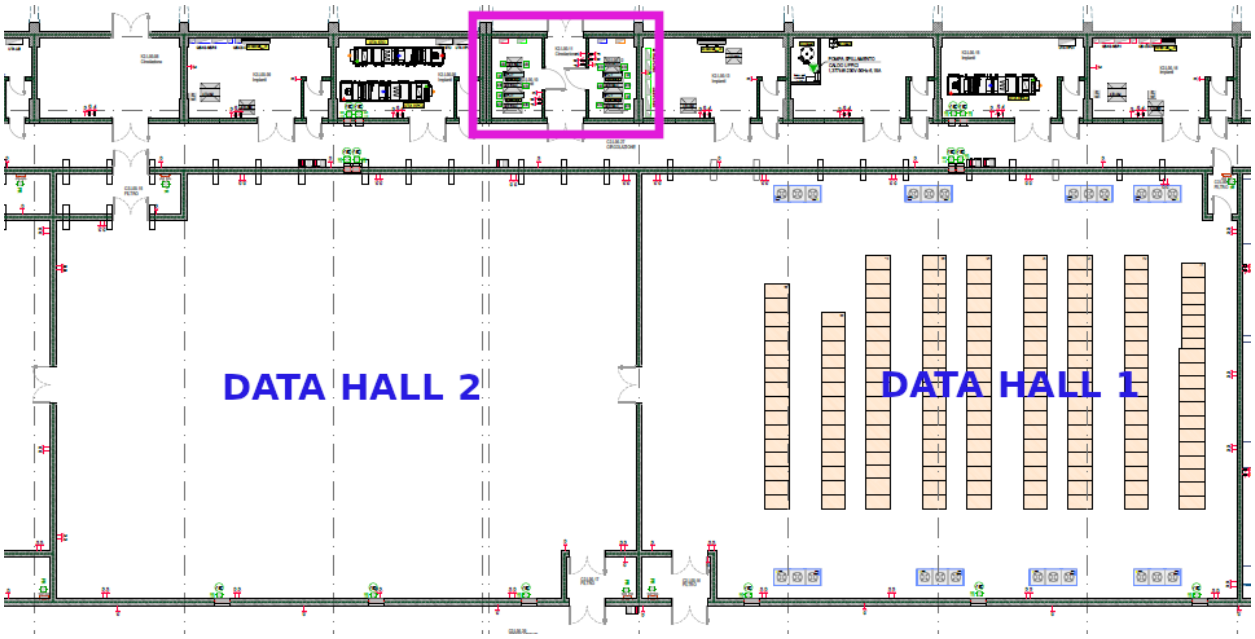

Figure 1. Technopole data centre floor plan. Leonardo layout is depicted in Data Hall 1 (on the right) and the procured System will be hosted in Data Hall 2 (on the left). The network room is indicated by a purple box at the top of the figure.

4.2 Data centre specifications

The floor space planimetry of the Data Hall 2 is reported in Figure 2. The tile size is 600x600 mm. In the red rectangle is shown the expected location of the procured solution.

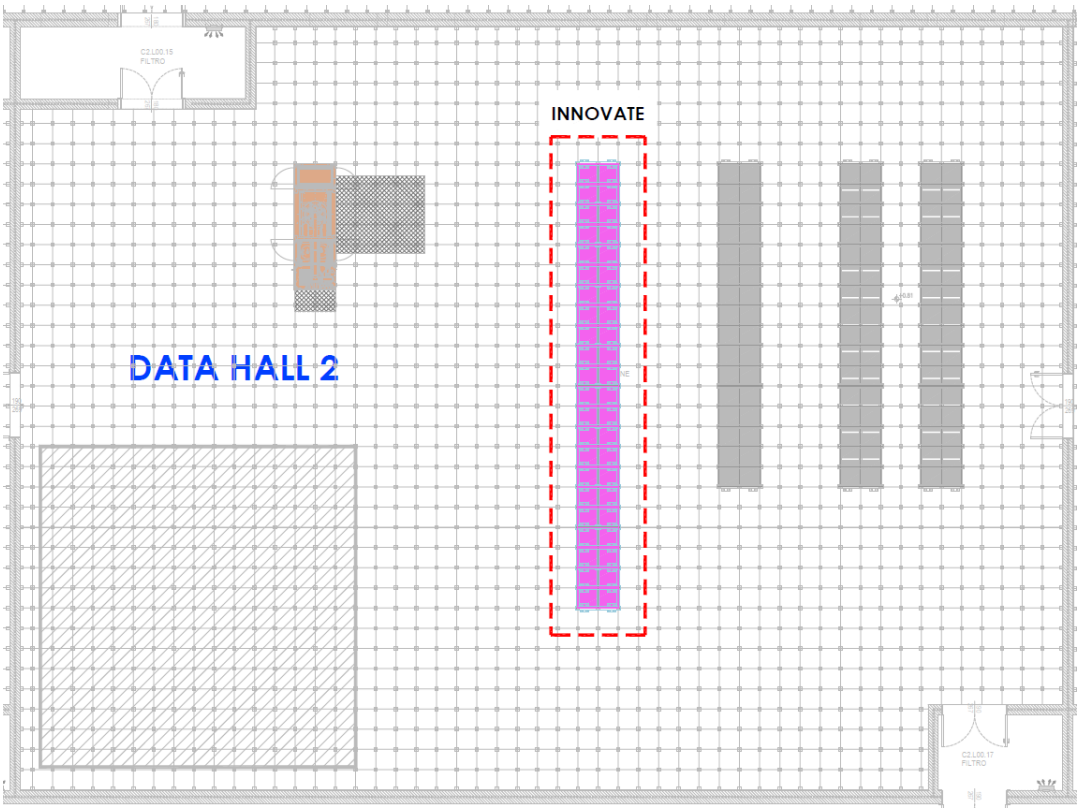

Figure 2. Floor plan of Data Hall 2 that should host the computing partitions of the infrastructure. The red rectangle provides just an indication for the whitespace available.

For the Data Hall 2 the following specifications apply:

Req.	Description	Category
4.2-1	*Dedicated data hall* The offered solution infrastructure will be installed the Data Hall 2. In terms of whitespace available, the Data Hall has no pillars and hosts a multitude of other systems as reported in Figure 2. Surface available for the offered solution is in the order of 40 m^2.	DCS
4.2-2	*Raised floor details* The data hall is equipped with a raised floor with height of 100 cm.	DCS
4.2-3	*Raised floor load* The raised floor will be reinforced to host DLC racks. The maximum expected load is 30 kN/m^2 and 11 kN single point load.	DCS

4.2-4	*Rack maximum height* The cable tray is at 240 cm of height.	DCS
4.2-5	*Room humidity* The offered infrastructure must comply with a relative humidity in the interval 20-60%.	DCS

4.2.1 Electrical infrastructure

Req.	**Description**	**Category**
4.2.1-1	*Data hall power supply* The Data Hall 2 has an electrical power supply with a frequency of 50 Hz, 3 x 400 Vac between the power lines, 230 Vac between the line and neutral. The hall can supply a power dedicated to the procured HPC infrastructure up to a maximum of 600 kW IT during the acceptance phase testing, and 400 kW IT in operating conditions.	DCS
4.2.1-2	*Electric load layout* The IT load installed in the Data Hall 2 is available through 1 rack row. The rack will be connected to 2 dedicated busbars installed above the racks. The computing racks will be connected to the power supplies directly with cables, i.e., without the use of plugs and without compromising the guarantee of the offer's components. The total power distributed to the racks cannot in any case exceed the limits defined in Req. 4.2.1-1.	DCS

4.2.2 Cooling infrastructure

The cooling of the procured infrastructure will rely on the cooling infrastructure already built and in use for Leonardo and other systems. Any change in the cooling water temperature set points will impact Leonardo operations – including its operating costs.

Req.	Description	Category
4.2.2-1	*Cooling infrastructure* The cooling infrastructure of the data centre produces tempered water (36°C) and chilled water (19°C). The tempered water is dedicated to the DLC Compute Nodes and the chilled water is used for the air conditioning of the data halls. The procured infrastructure needs to comply with this cooling infrastructure requirement and the limits reported in req. 4.2.2-2 and 4.2.2-3.	DCS

4.2.2-2	*Liquid cooling* Tempered water will be used for direct cooling of the system. The tempered water circuit is at 36°C inlet. The tempered water circuit has a cooling capacity of 600 kWf dedicated for the procured infrastructure.	DCS
4.2.2-3	*Air cooling* The air-cooling system of the Data Hall 2 is currently based on 4 CRAHs. Each device has a cooling power of roughly 100 kWf (400kWf aggregated). 2 more CRAHs will be installed in the Data Hall 2 with the same specific as the existing ones giving the procured system a dedicated air-cooling capacity of roughly 200 kWf.	DCS
4.2.2-4	*Flow rate* The maximum tempered water flow rate available for procured system is 900 l/min.	DCS

4.3 Data Hall MEP layout

Power and mechanical distribution and their arrangement in the Data Hall 2 are reported in the figures below.

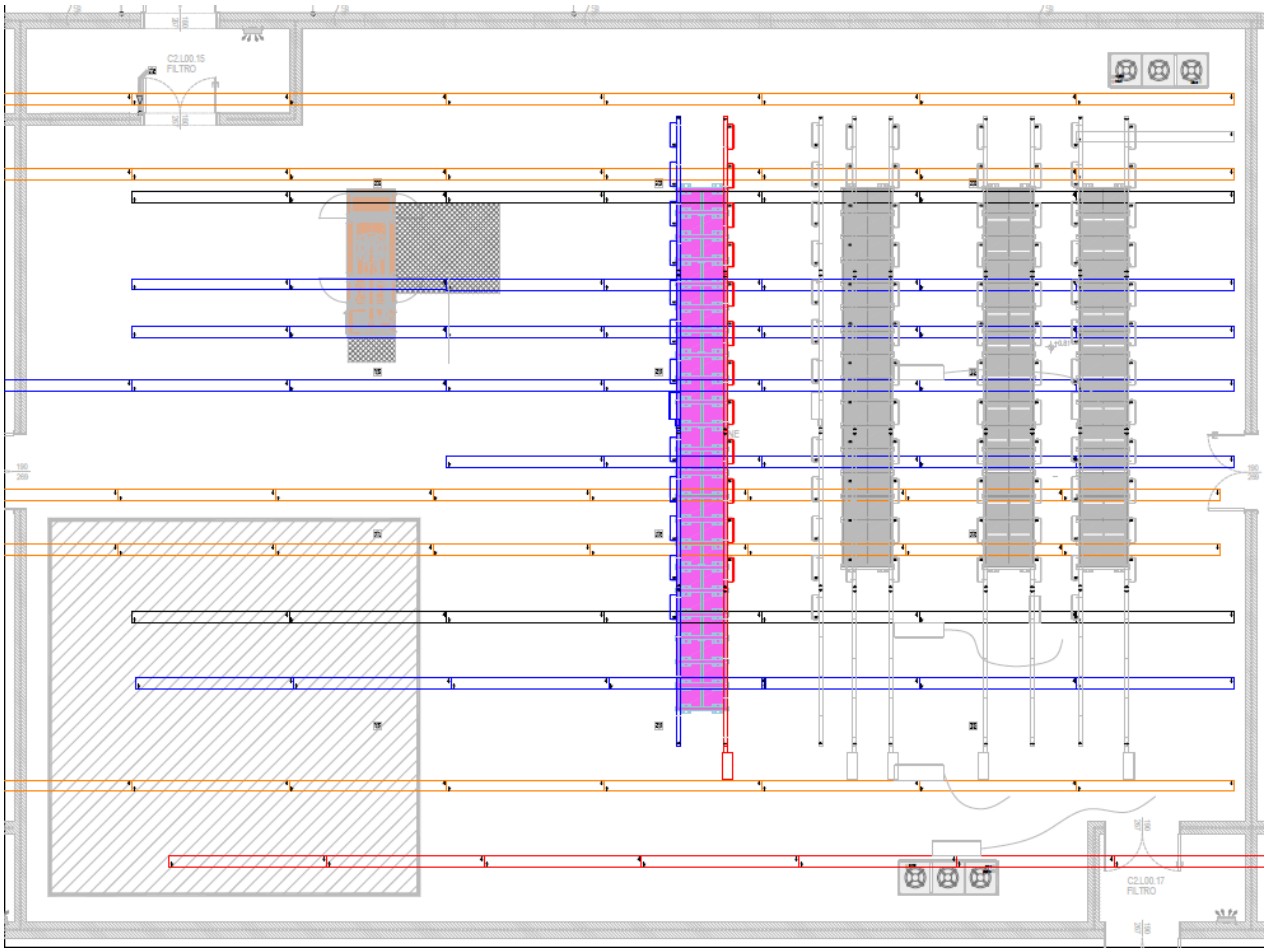

Figure 3. Power distribution layout of Data Hall 2. Each colour represents one of 4 electric branches.

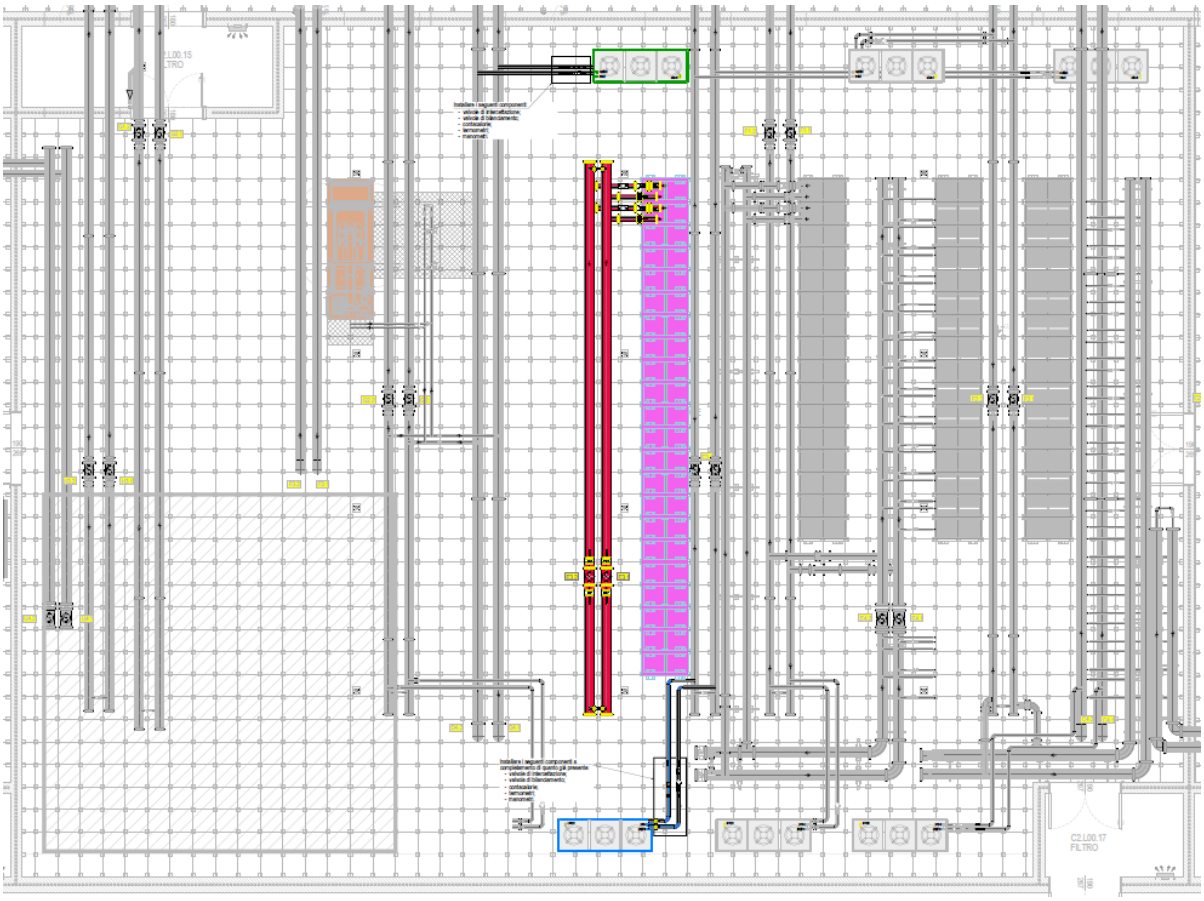

Figure 4. Cooling distribution layout of data Hall 2.

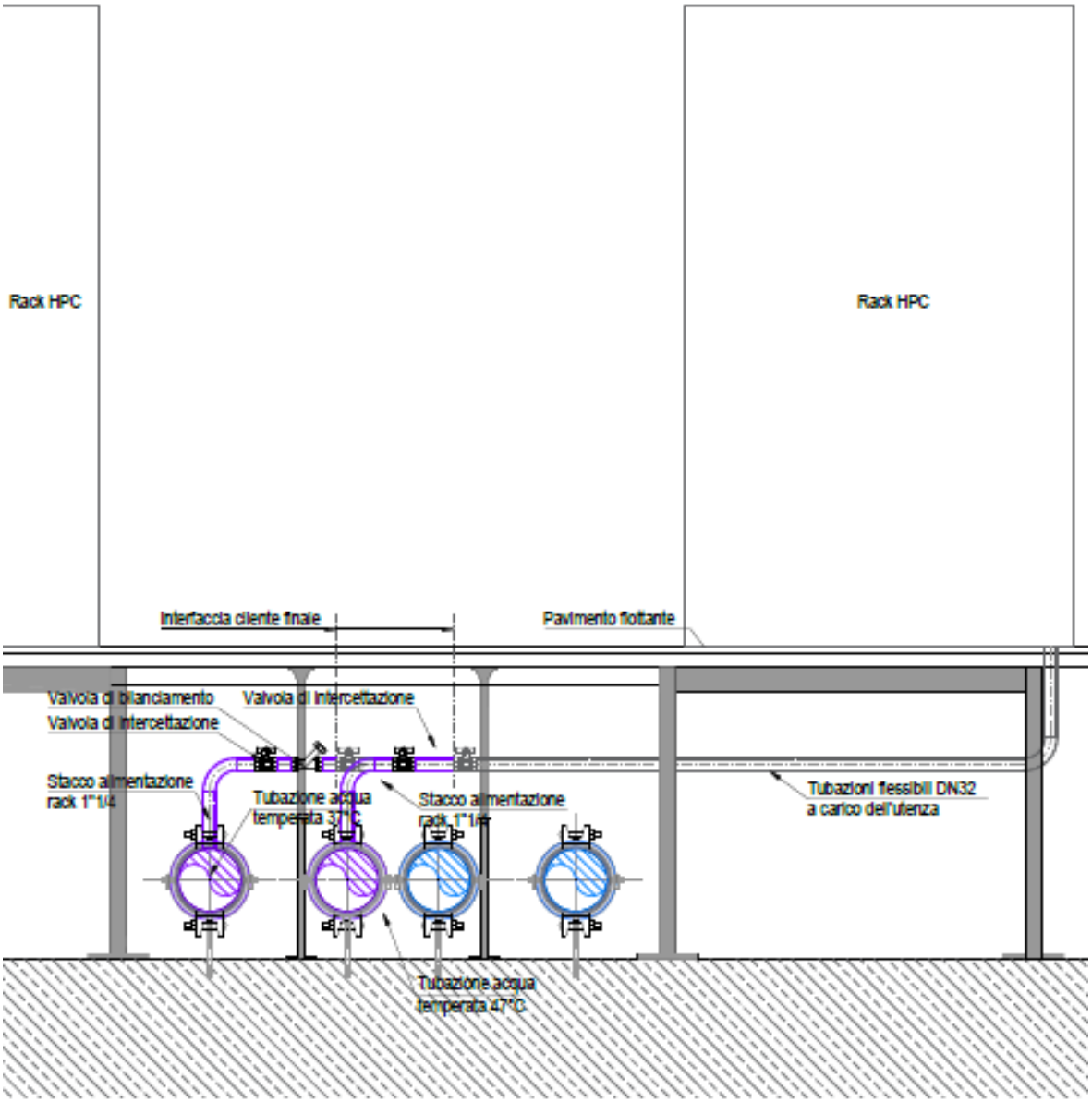

Figure 5. Section of the cooling distribution under the raised floor.

5 System infrastructure

5.1 General requirements

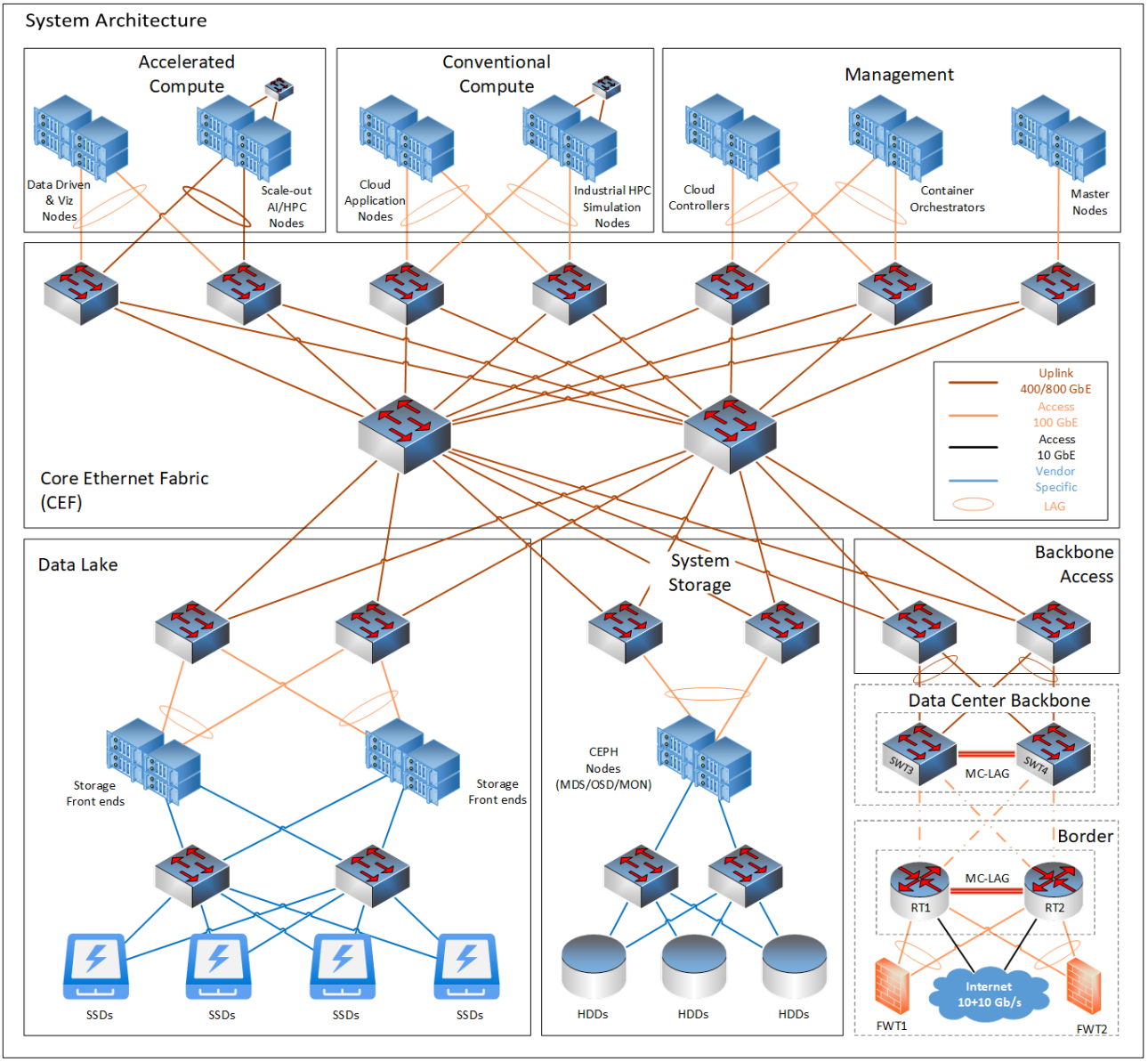

Figure 6: Reference design of the system architecture.

5.1.1 Functional aspects

This section outlines the requirements for a complete and fully integrated platform capable of delivering reliable service operation and centralized system management. The infrastructure must support automated provisioning, remote supervision, fault monitoring, and maintenance procedures designed to minimize service disruption, while ensuring full visibility into system health and performance.

Req.	Description	Category
5.1.1-1	*Integrated platform*	MRQ

	The infrastructure shall be designed as a comprehensive and fully integrated platform. The Proposal shall include all necessary hardware and software components to deliver user-facing services and to ensure complete operational management of the system. The solution shall be fully operational upon deployment and must support scalability, reliability, and maintainability in alignment with defined operational and performance objectives.	
5.1.1-2	*Reboot time* Each system partition shall support a full reboot within a maximum duration of 60 minutes. This requirement guarantees timely restoration of functionality and operational continuity following maintenance activities, updates, or unforeseen disruptions. The solution must incorporate efficient boot mechanisms, optimized initialization sequences, and resilient fault recovery processes to meet this performance criterion.	MRQ
5.1.1-3	*Common node features* Each node shall be equipped with a Board Management Controller (BMC) and remote monitoring functionalities meeting the following minimum specifications: Board Management Controller: • Provision of a dedicated or shared Ethernet network interface for management purposes. • Support for remote management via web-based interfaces, such as Java and/or HTML5 GUI. • Availability of virtual console and virtual media (VMedia) functionalities. • System lockdown functionality to prevent unintended configuration changes during operation. When enabled, all configuration modifications, including firmware updates, shall be blocked and the user shall be notified accordingly. • Support for digitally signed firmware updates to ensure integrity and authenticity. • Firmware rollback capabilities to restore previous stable versions. • Implementation of protection mechanisms for firmware updates of internal components. • Enforcement of secure default password practices. • Support for centralized authentication via LDAP. • Capability to block IP addresses as a security measure. Remote Monitoring and Alerting: • The system shall be capable of automatically issuing alerts to the vendor's support service, including all relevant diagnostic information, without requiring intervention from system administrators. • For critical components such as RAM and disk drives, the system shall automatically detect and report pre-failure conditions to the support service.	MRQ
5.1.1-4	*Integrated cluster management system* The solution shall be provided with a fully integrated software management stack for the centralized administration of all cluster resources, including node	MRQ

	provisioning, basic hardware monitoring, and operating system supervision. The proposed software shall support comprehensive out-of-band management of all hardware components and facilitate the automated provisioning and reconfiguration of system nodes. The solution shall enable the automation of all essential system management operations, including but not limited to: • Installation of the operating system on compute and Management Nodes. • Reconfiguration of the operating system on individual nodes or, where applicable, across all managed infrastructure components. • Collection of diagnostic and health information from nodes and, if supported, from all integrated hardware apparatus. • Firmware update and version management for all nodes. This management software shall typically run on the Management Partition and shall support a redundant (high availability) configuration to ensure uninterrupted operation of critical management functions.	
5.1.1-5	*Health Monitoring, Validation, and Automated Recovery* The solution shall be equipped with a comprehensive set of tools to perform health checks and configuration validation of all hardware and software components within the system. These tools shall be designed for integration with the workload manager, enabling automated enforcement of job scheduling policies that utilize only fully functional and validated components. Where applicable, the provided tools shall support automated recovery actions, such as node isolation, service restart, or hardware reset, to remediate detected faults. All health check outcomes and recovery actions shall be fully logged and made accessible for audit, diagnostics, and system administration purposes.	MRQ
5.1.1-6	*Rolling Update and Maintenance Mechanisms* The system shall support rolling update mechanisms that enable the execution of reliable software updates and selected maintenance operations with minimal accumulated downtime and without impacting the availability of critical services or running workloads. During full system maintenance, the infrastructure shall implement a staged idle-time strategy, wherein Compute Nodes progressively transition to an idle state as active jobs complete. No new jobs shall be scheduled on nodes reserved for upcoming maintenance, thereby facilitating a graceful and workload-aware shutdown. This capability shall substantially reduce maintenance overhead and improve system availability by avoiding abrupt job termination and maximizing resource utilization up to the initiation of maintenance procedures.	MRQ
5.1.1-7	*Cluster Hardware Event Monitoring* The cluster management software shall support both out-of-band and, where applicable, in-band monitoring of hardware events, including but not limited to system event logs, machine check exceptions, and other platform-reported fault or alert conditions. All detected events shall be collected in near real time and stored in a centralized repository, ensuring consistent availability for system administrators, diagnostic	MRQ

	tools, and audit processes. The solution shall facilitate timely analysis and correlation of hardware-related issues across the infrastructure to support proactive maintenance and operational reliability.	
5.1.1-8	*Monitoring infrastructure* The system shall include a comprehensive hardware and software framework for monitoring the health status of all system components. The monitoring solution shall support the collection and reporting of health parameters at the component level and shall expose APIs and/or open-source frameworks to ensure seamless integration with widely adopted open-source monitoring tools. All hardware faults, including those with the potential to impact the performance or stability of Compute Nodes, interconnect devices, or any other system components, shall be automatically detected and reported by the monitoring infrastructure. The monitoring and management systems included in the infrastructure shall expose APIs that enable seamless integration with third-party monitoring and management frameworks. These APIs shall provide timely and comprehensive information on the operational status of all infrastructure components. The system shall ensure that any fault or degradation affecting system components is detected and made available through the APIs to support prompt alerting and incident response.	MRQ
5.1.1-9	*Node power and energy measurement* The infrastructure shall provide near real-time power and energy consumption measurements at the node level. The monitoring mechanism shall have minimal impact on system performance and scalability.	MRQ

5.2 Interconnects

This section defines the technical requirements for the interconnection networks that form the communication backbone of the infrastructure. It covers the Core Ethernet Fabric (CEF), High-Performance AI/HPC Fabric (HPAIF), HPC Simulation Fabric (HPCSF), and the Management Network, each designed to serve distinct traffic patterns and operational functions. These networks must ensure high throughput, low latency, scalability, and reliability to support the performance, monitoring, and management needs of the entire system.

5.2.1 Core Ethernet Fabric

The CEF is a dedicated interconnection network that facilitates data exchange among Compute Nodes, storage systems, and service components. It is engineered to support the consistent and high-throughput transfer of large volumes of data across the infrastructure. The CEF plays a pivoting role in ensuring fast and reliable access to persistent storage resources and in enabling seamless integration with the overall storage platform.

Req.	Description	Category

5.2.1-1	*General requirements* The supplied infrastructure shall include a dedicated fabric for storage and services communications (north-south traffic). The fabric shall meet the following technical requirements: • The technology employed shall be based on IP/Ethernet protocol. • The fabric shall be based on a minimum network technology of 800 Gbit/s with access links at 100/400 Gbit/s. The logical representation of the interconnection fabric is provided in Figure 6.	MRQ
5.2.1-2	*Bandwidth performance* The infrastructure shall include a network fabric that provides a maximum oversubscription of 2:1.	MRQ
5.2.1-3	*Topology* The infrastructure shall implement a leaf-spine Layer 3 architecture using BGP as the control plane with EVPN extensions, and VXLAN as the data plane. The design must provide sufficient radix and port density to support large-scale compute and storage workloads, ensuring scalability, multi-tenancy, and high availability.	MRQ
5.2.1-4	*Networking optimizations for storage communication* • The interconnection fabric shall implement hardware-based congestion detection and mitigation mechanisms, including Priority Flow Control (PFC) and Explicit Congestion Notification (ECN), with support for end-to-end congestion management mechanisms such as DCQCN or equivalent standards-based approaches, to ensure lossless and stable data transfers for storage workloads. • The interconnection fabric shall be explicitly optimized for high-throughput and deterministic performance, supporting communication patterns characteristic of NVMe-based storage systems, including full support for RDMA over Converged Ethernet (RoCE v1/v2), with RoCE v2 support required for routed spine-leaf topologies. • The fabric shall support NVMe over Fabrics (NVMe-oF) over Ethernet, including compatibility with NVMe multipathing and ECMP-based traffic distribution. • Support for deterministic ECMP forwarding and in-order packet delivery for RDMA and NVMe-oF traffic shall be provided. • Advanced buffer management and per-priority traffic isolation mechanisms shall be available to prevent head-of-line blocking and congestion propagation. • The fabric shall support real-time telemetry and monitoring of storage-related traffic, including PFC events, ECN marking, queue depth, and latency metrics. • Support for fast path failure detection (e.g. BFD) to minimize storage I/O disruption.	MRQ
5.2.1-5	*Network capabilities*	MRQ

	The switches comprising the Core Ethernet Fabric shall support the following network functionalities: • Support for active-active multi-homing using a multi-chassis link aggregation protocol, such as MC-LAG, MLAG, or an equivalent standards-based or de-facto interoperable technology. • Support for IEEE 802.3ad / 802.1AX Link Aggregation (LACP), including operation across multi-chassis configurations. • Support for VLANs and VLAN tagging compliant with IEEE 802.1Q, including VLAN trunking. • Support for BGP (Border Gateway Protocol) for IP fabric and ECMP routing in spine-leaf architecture. • Support for VXLAN (Virtual Extensible LAN) as the primary overlay encapsulation technology. • Support for VXLAN routing through an integrated control plane, such as EVPN (Ethernet VPN) based on BGP, including MAC learning and IP prefix advertisement. • Support for VXLAN bridging and routing (L2 and L3 VNIs). • Support for fast control-plane convergence and sub-second failure recovery in case of link or node failures. • Support for graceful restart and non-disruptive upgrades where applicable (e.g. BGP graceful restart). • Support for Quality of Service (QoS) features, including: o Traffic classification and marking (IEEE 802.1p / DSCP). o Priority queuing and congestion management. • The interconnection fabric should provide full support to IPv4 and IPv6.	
5.2.1-6	*Participants* All nodes shall be connected to the fabric as shown in Figure 6.	MRQ
5.2.1-7	*Border leaf* The Data Fabric shall be provisioned with one or more border leaf switches, that shall be connected to the CINECA backbone, which is based on NVIDIA Spectrum SN4700 named as SWT3 and SWT4 in Figure 6. These border leaf switches shall provide high-throughput connectivity between the CEF and the data centre backbone. The architecture shall ensure that data routing through these connections shall provide optimal performance and fault tolerance. All necessary cabling, transceivers, and related accessories required to implement and operate these high-speed links shall be included in the scope of supply. The Supplier shall also be responsible for the complete execution of cable laying and physical installation activities required to establish the connection between the border leaf switches and the data centre backbone. Integration and configuration activities concerning the interconnection with the data centre backbone shall be carried out in close coordination with CINECA technical staff to ensure full compatibility, proper operation, and seamless integration within the existing infrastructure.	MRQ
5.2.1-8	*Monitoring and managing capabilities.*	MRQ

	The following capabilities shall be provided as part of the interconnection fabric: • The switching platform shall provide built-in capabilities for near real-time telemetry and monitoring, including queue depth, port utilization, packet drops, congestion events, and flow-level statistics. Telemetry data shall be accessible via standard or open APIs (e.g., gNMI, REST, sFlow, or equivalent) to facilitating proactive monitoring and diagnostics. • Each fabric switch shall be equipped with a dedicated out-of-band management port.	
5.2.1-9	*Power redundancy* Each fabric switch shall be equipped with redundant and hot-swappable power supply units to ensure high availability and maintainability without requiring system downtime.	MRQ

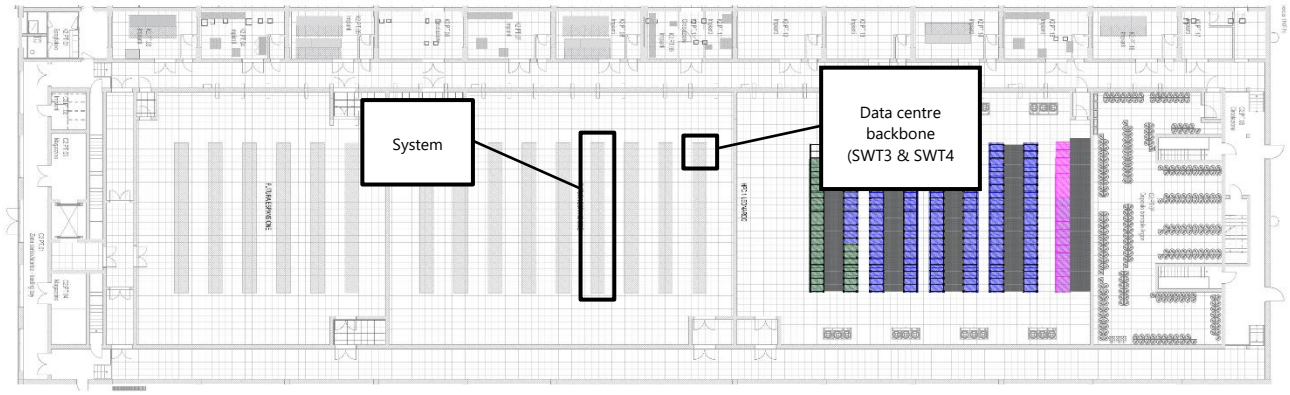

Figure 7: Location of the data centre backbone (SWT3 & SWT4 switch)

5.2.2 High-Performance AI Fabric

The High-Performance AI Fabric (HPAIF) is a high-performance interconnect network specifically designed to support GPU-accelerated workloads. It enables low-latency, high-bandwidth communication across the scale-out AI/HPC partition, ensuring efficient data exchange between GPUs distributed throughout the infrastructure. This fabric is dedicated to support large-scale AI operations, supporting intensive inter-node communication patterns typical of training and inference workflows at scale.

Req.	Description	Category
5.2.2-1	*General requirements* The infrastructure shall include a high-performance fabric dedicated for GPU-to-GPU communication (east-west traffic) of AI/HPC Compute Nodes. The fabric shall meet the following technical requirements: • The fabric shall be based on InfiniBand, Spectrum-X, Ultra Ethernet (only mandatory requirements), or standard Ethernet technology. • If the fabric is based on standard Ethernet technology, it shall be implemented as a single-switch fabric. In such case, the Ethernet fabric shall consist of either a single switch or a single modular chassis switch, in order to ensure single-hop latency and avoid inter-switch forwarding. • The fabric shall be based on a minimum network technology of 400 Gbit/s.	MRQ

5.2.2-2	*Bandwidth performance* The network fabric shall provide a maximum oversubscription ratio of 1:1 (no-blocking), ensuring high throughput and balanced traffic distribution across system components while allowing for a controlled level of resource contention. This configuration is intended to optimize infrastructure cost and scalability without significantly compromising application performance. The selected oversubscription ratio shall be supported by appropriate congestion management mechanisms and traffic engineering capabilities to maintain predictable and stable performance under typical and peak workload conditions.	MRQ
5.2.2-3	*Topology* The switching infrastructure shall support scalable topologies such as Fat-tree, or leaf-spine architectures, with sufficient radix and port density to support large-scale AI workloads.	MRQ
5.2.2-4	*AI/HPC optimizations* • The fabric switches shall implement hardware-based congestion detection and mitigation mechanisms, including support for credit-based flow control or equivalent lossless transport techniques. These mechanisms must enable reliable operation of remote direct memory access (RDMA) protocols under sustained high-throughput conditions. • The switching infrastructure shall support dynamic traffic optimization features such as adaptive routing, multipath load distribution, and telemetry-driven congestion avoidance. These capabilities must operate at the fabric level to ensure consistent performance across large-scale, low-latency network environments. • The fabric shall be architected to support deterministic, lossless communication patterns with low latency and high throughput, suitable for distributed AI/ML training. Compatibility with collective communication acceleration frameworks and native RDMA transport is required. • The platform shall implement low-latency switching techniques such as cut-through or partial cut-through forwarding, and support adaptive path selection mechanisms that respond dynamically to congestion or imbalance across the fabric. • The fabric shall be designed for horizontal scalability, supporting deployments with hundreds of compute accelerators and capable of handling increasing volumes of data and inter-node traffic without degradation in performance.	TRQ
5.2.2-5	*Participants* Only Scale-out AI/HPC and Management Nodes shall be connected to the HPAIF.	MRQ
5.2.2-6	*Monitoring and managing capabilities.* The following capabilities shall be provided as part of the fabric: • The switching platform shall provide built-in capabilities for real-time telemetry and monitoring, including queue depth, port utilization, packet drops, congestion events, and flow-level statistics. Telemetry data shall be	MRQ

	accessible via standard or open APIs (e.g., gNMI, REST, sFlow, or equivalent) to facilitating proactive monitoring and diagnostics. • Each fabric switch shall be equipped with a dedicated out-of-band management port.	
5.2.2-7	*Power redundancy* Each fabric switch shall be equipped with redundant and hot-swappable power supply units to ensure high availability and maintainability without requiring system downtime.	MRQ

5.2.3 HPC Simulation Fabric

The HPC Simulation Fabric (HPCSF) is a high-performance network specifically engineered to support CPU-based workloads in industrial high-performance computing environments. It enables ultra-low latency and high-bandwidth communication across Compute Nodes, ensuring efficient data exchange and synchronization for distributed memory applications. This fabric is optimized for tightly coupled, communication-intensive workloads typical of industrial simulations, numerical modeling, and large-scale data processing, delivering consistent performance and scalability across complex HPC infrastructures.

Req.	Description	Category
5.2.3-1	*General requirements* The infrastructure shall include a high-performance scale-out fabric dedicated for CPU-to-CPU communication (east-west traffic) of HPC Compute Nodes. The fabric shall meet the following technical requirements: • The fabric shall be based on InfiniBand technology. • The fabric shall be based on a minimum network technology of 400 Gbit/s.	MRQ
5.2.3-2	*Bandwidth performance* The network fabric shall provide a maximum oversubscription ratio of 1:1, ensuring high throughput and balanced traffic distribution across system components while allowing for a controlled level of resource contention. This configuration is intended to optimize infrastructure cost and scalability without significantly compromising application performance. The selected oversubscription ratio shall be supported by appropriate congestion management mechanisms and traffic engineering capabilities to maintain predictable and stable performance under typical and peak workload conditions.	MRQ
5.2.3-3	*Topology* The switching infrastructure shall support scalable topologies such as Fat-tree, Dragonfly(+), folded Clos, or leaf-spine architectures, with sufficient radix and port density to support large-scale HPC workloads.	MRQ
5.2.3-4	*HPC optimizations* • The interconnection fabric shall be based on InfiniBand or equivalent technology and shall support InfiniBand NDR 400 Gb/s per port, providing a lossless, low-latency transport optimized for large-scale HPC workloads. The fabric shall implement native InfiniBand credit-based flow control mechanisms ensuring zero packet loss, deterministic latency, and reliable operation of RDMA under sustained high-throughput conditions. • The switching infrastructure shall natively support InfiniBand fabric services, including adaptive routing, multipath traffic distribution, and congestion-aware forwarding. These capabilities shall operate at the fabric level, leveraging InfiniBand routing and congestion control mechanisms to maintain consistent performance across large-scale, low-latency deployments. • The fabric shall be architected to support deterministic, lossless communication patterns with ultra-low end-to-end latency and high bandwidth efficiency, suitable for tightly coupled HPC simulations workloads. Full compatibility with native InfiniBand RDMA transports and collective communication acceleration frameworks (such as hardware-assisted collective offload) is required to efficiently support parallel applications.	MRQ

	• The fabric shall support in-network computing capabilities, enabling hardware-assisted reduction and aggregation operations to be executed directly within the InfiniBand fabric. These capabilities shall accelerate collective communication patterns (e.g. all-reduce, broadcast, barrier operations), reducing end-to-end latency, network congestion, and host CPU utilization for large-scale MPI workloads. • The switch platform shall implement ultra-low-latency forwarding techniques inherent to InfiniBand switching architectures and shall support adaptive path selection mechanisms that dynamically respond to congestion, link imbalance, or fabric events, ensuring optimal traffic distribution and sustained performance. • The fabric shall support operation under a vendor-supported InfiniBand network operating system and shall integrate with a standards-compliant InfiniBand Subnet Manager. Integration with external fabric management and monitoring tools shall be supported to enable centralized orchestration, diagnostics, and fine-grained telemetry of large-scale InfiniBand fabrics. • The fabric shall be designed for horizontal scalability, supporting non-blocking InfiniBand topologies with high-radix switching elements. The architecture shall support deployments with hundreds of compute nodes and accelerators, while maintaining consistent bandwidth, latency, and collective communication performance at scale.	
5.2.3-5	*Participants* Only HPC Simulation and Management Nodes shall be connected to the HPCSF.	MRQ
5.2.3-6	*Monitoring and managing capabilities.* The following capabilities shall be provided as part of the fabric: • The switching platform shall provide built-in capabilities for real-time telemetry and monitoring, including queue depth, port utilization, packet drops, congestion events, and flow-level statistics. Telemetry data shall be accessible via standard or open APIs (e.g., gNMI, REST, sFlow, or equivalent) to facilitating proactive monitoring and diagnostics. • Each fabric switch shall be equipped with a dedicated out-of-band management port.	MRQ
5.2.3-7	*Power redundancy* Each fabric switch shall be equipped with redundant and hot-swappable power supply units to ensure high availability and maintainability without requiring system downtime.	MRQ

Management network

The Management Network provides the communication backbone for system administration, monitoring, and control across all infrastructure components. It supports both in-band and out-of-band management functions, allowing for secure and reliable operations, including system provisioning, diagnostics, and maintenance. The network is designed to ensure high availability and operational isolation, forming the foundation for efficient lifecycle and configuration management of the system.

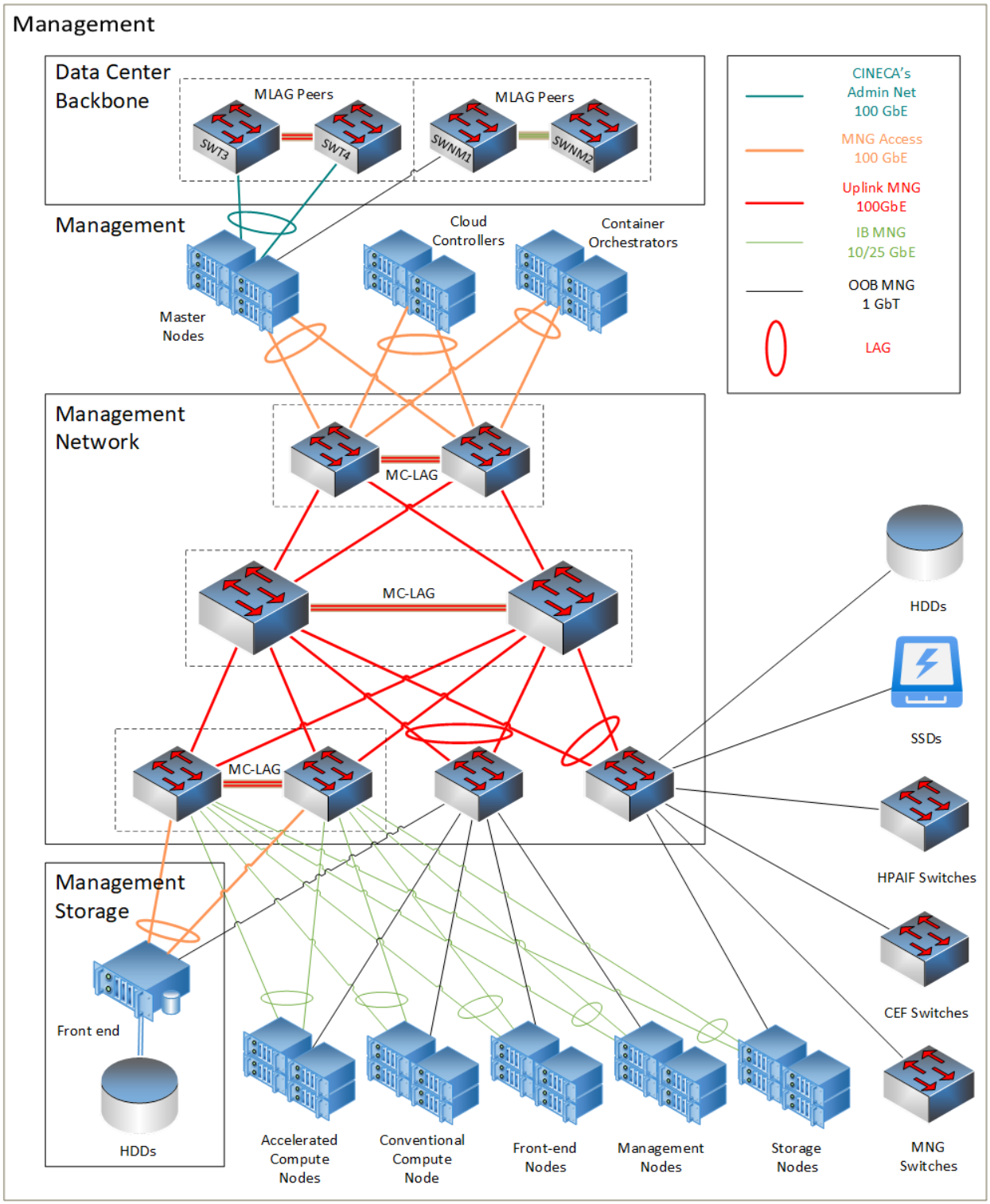

Figure 8: Reference design of the management network.

Req.	Description	Category
5.2.4-1	*Management Network* The system shall include a physically dedicated Ethernet network for management purposes, meeting the following minimum technical requirements: • A leaf-spine (L3) or access/aggregated (L2) architecture to ensure high availability and operational flexibility. • The maximum allowable oversubscription ratio for the Management Network topology shall not exceed 4:1. • The network design shall incorporate full redundancy, so the failure of a single switch or link shall not impact the stability or continuity of the Management Network. • The Management Network shall support the configuration of at least two logically separated VLANs to implement both in-band and out-of-band management functions. The structure of the Management Network is illustrated in Figure 8.	MRQ
5.2.4-2	*Network Participants* • *In-Band (IB) Management Network*: This logical network shall be utilized for the management and deployment of Compute Nodes supporting operational services such as bare-metal or operating system installation, system monitoring, and metering. • *Out-of-Band (OOB) Management Network*: This logical network shall be used for the management of the BMC interfaces of all infrastructure components, including but not limited to storage systems, network equipment, and chassis.	MRQ
5.2.4-3	*Monitoring and managing capabilities* The following capabilities shall be provided as part of the Management Network: • It shall support mechanisms for system management and near real-time collection of performance and health monitoring data. Protocols such as sFlow or equivalent shall be supported to enable continuous telemetry and diagnostics. • Each switch shall be equipped with a dedicated OOB management port, which shall be connected to the OOB Management Network.	MRQ
5.2.4-4	*Aggregation switches* • Aggregation switches shall provide a sufficient number of ports to ensure connectivity with the access switches. • Aggregation switches shall support both static and dynamic Layer 3 routing functionalities, including but not limited to protocols such as static routing, OSPF, BGP, MP-BGP, and OSPFv3.	MRQ
5.2.4-5	*Access switches* The Management Network shall meet the following minimum port configuration requirements:	MRQ

	• It shall provide access ports operating at 1 Gb/s (Base-T) for the connection OOB management interfaces. • It shall provide access ports operating at 10/25 GbE for the connection of IB management interfaces. • It shall provide access ports operating at 25/100 GbE for the connection of Management Nodes and storage systems.	
5.2.4-6	*Network capabilities* The switches comprising the Core Ethernet Fabric shall support the following network functionalities: • Support for active-active multi-homing using a multi-chassis link aggregation protocol (MC-LAG or an equivalent technology). • Support for Virtual LANs (VLAN), Virtual Extensible LAN (VXLAN), and compliance with widely adopted IEEE 802 standards, including but not limited to 802.1Q (VLAN tagging), 802.1p (QoS), and 802.3ad (Link Aggregation).	MRQ
5.2.4-7	*Network performance* The Management Network shall meet the following performance requirements: • The network performance shall enable the full reconfiguration of the operating system (without requiring reinstallation) on all Compute Nodes. • The network shall support the (re-)installation of the operating system on all Compute Nodes. • The network shall support the cold reboot of all Compute Nodes. • The network shall allow for the complete collection of metrics and sensor data from the management interfaces (e.g., BMCs) of all Compute Nodes and network devices using vendor-native or open-standard tools and protocols (such as IPMI tool, Redfish, Confluent, SNMP, etc.), leveraging as many parallel sessions as supported by the monitoring infrastructure.	MRQ
5.2.4-8	*Power redundancy* Each fabric switch shall be equipped with redundant and hot-swappable power supply units to ensure high availability and maintainability without requiring system downtime.	MRQ

5.3 Accelerated Compute Partition

The Accelerated Compute partition is designed to host high-performance computing nodes equipped with hardware accelerators (GPUs). These nodes are intended to support workloads that require massive parallelism and high-throughput data processing. The infrastructure must ensure optimal performance for AI training and scientific simulations by leveraging acceleration frameworks.

5.3.1 Cloud data driven & visualization Sub-Partition

The Cloud data driven & visualization sub-partition supports data-centric workloads, cloud-native applications, AI inference tasks, and visualization workflows. These nodes are optimized for high throughput, flexible orchestration, and integration with storage and data platforms. The infrastructure enables scalable deployment

of services handling large volumes of structured and unstructured data and supports interactive visualization such as remote rendering and real-time data exploration. The following requirements define the technical capabilities expected from this sub-partition.

Req.	Description	Category
5.3.1-1	*Partition size* The Cloud data driven & visualization sub-partition must include eight (8) nodes.	MRQ
5.3.1-2	*Common node requirements* The nodes shall implement the requirements provided in Req. 5.1.1-3.	MRQ
5.3.1-3	*Node configuration* Each node shall be equipped with the following components: • Two (2) state-of-the-art, server-grade CPUs. The node must provide a minimum of thirty-two (32) physical CPU cores for every GPU installed within the node. • Eight (8) high-end graphics processing units (GPUs) optimized for AI inference tasks and professional visualization.	MRQ
5.3.1-4	*GPU Technology* The proposed system shall integrate state-of-the-art GPU accelerators explicitly designed to support high-performance AI inference and visualization workloads. Each GPU shall meet the following minimum technical specifications: • Support hardware-level optimization for AI inference and training, including native capabilities for tensor operations and mixed-precision arithmetic (e.g., FP/BF16, FP8, FP4), with particular applicability to inference LLMs. • Be equipped with a minimum of 96 GB of GDDR7 (or HBM) with ECC per GPU, ensuring adequate memory bandwidth and capacity to inference large-scale deep learning models efficiently. • Support IEEE-conforming single precision (FP32) computation for accelerated inference and provide 3D hardware acceleration via OpenGL or equivalent industry-standard graphics APIs. The configuration shall ensure the computational and rendering capabilities required for both large-scale inference pipelines and advanced remote visualization, including interactive sessions and real-time data exploration.	MRQ
5.3.1-5	CPU Technology The CPUs shall be based on the x86_64 architecture and support hardware-assisted virtualization and containerization technologies (e.g., Intel VT-x/AMD-V, VT-d/IOMMU), as well as confidential computing capabilities (e.g., Intel TDX, AMD SEV) to enable secure execution of sensitive workloads in isolated environments.	
5.3.1-6	*System memory configuration* Each node shall be equipped with a minimum of 768 GB of system memory. Memory shall be implemented using DDR5, LP-DDR5, or newer technologies, and configured	MRQ

	to maximize effective memory bandwidth, channel utilization, and overall system reliability in order to guarantee sustained performance for memory-intensive computational workloads.	
5.3.1-7	*Network requirements* Each node shall be equipped with the following network interfaces to ensure high-performance connectivity, system management, and network function offloading capabilities: CEF Connectivity: • One (1) NIC with two (2) Ethernet ports, each delivering a minimum of 100 Gb/s, for an aggregated throughput of 200 Gb/s. The two network ports shall be connected to distinct top-of-rack or leaf switches to ensure path redundancy and fault tolerance. The NIC shall support the following hardware capabilities: o VLAN offloading in hardware. o VXLAN encapsulation and decapsulation offloading in hardware. o RoCE versions 1 and 2. o RoCE support over overlay networks (VXLAN/NVGRE). o NVMe/TCP target offload capabilities. o TCP/UDP/IP stateless offloading. o SR-IOV. IB Management Connectivity: • One (1) NIC featuring one (1) Ethernet port operating at 10/25 Gb/s. The NIC shall support remote boot over Ethernet leveraging industry-standard mechanisms such as PXE, HTTP Boot (UEFI-based), NVMe/TCP Boot. OOB Management Connectivity: • One (1) NIC featuring a single (1) Ethernet port operating at 1 Gb/s (Base-T) providing access to the OOB Management Network via the BMC. Note: The OOB and IB management interfaces may share the same physical port, provided that all port connectivity and functional requirements are fully met.	MRQ
5.3.1-8	*Local storage* Each node shall be equipped with a RAID 1 (mirroring) configuration using SSD drives, ensuring redundancy and providing a minimum of 200 GB of usable capacity, dedicated to the operating system and critical system services.	MRQ

5.3.2 Scale-out AI/HPC Sub-Partition

The Scale-out AI/HPC Sub-Partition is designed for distributed, high-performance workloads typical of industrial AI and scientific computing. These nodes support scale-out execution models, enabling efficient parallel processing across large numbers of compute resources. Interconnected via a high-bandwidth, low-latency fabric (HPAIF), the infrastructure ensures reliable inter-node communication, collective operations, and consistent performance for demanding AI training and simulation tasks. The following requirements define the essential capabilities of this partition.

Req.	Description	Category
5.3.2-1	*Partition size* The Scale-out AI/HPC sub-partition shall comprise, at minimum, either sixteen (16) nodes equipped with eight (8) GPUs each, or thirty-two (32) nodes equipped with four (4) GPUs each.	MRQ
5.3.2-2	*Common requirements* The nodes shall implement the requirements provided in Req. 5.1.1-3.	MRQ
5.3.2-3	*Cooling technology* The Scale-out AI/HPC platforms shall employ Direct Liquid Cooling (DLC) solutions capable of dissipating at least 65% of the system's total TDP through the liquid cooling loop. The DLC implementation shall be designed to ensure thermal stability under sustained high-load operation, minimize reliance on air cooling, and enable efficient heat recovery strategies.	MRQ
5.3.2-4	*Target cooling technology* The Scale-out AI/HPC platforms should employ Direct Liquid Cooling (DLC) solutions capable of dissipating at least 90% of the system's total TDP through the liquid cooling loop. The DLC implementation should be designed to ensure thermal stability under sustained high-load operation, minimize reliance on air cooling, and enable efficient heat recovery strategies.	TRQ
5.3.2-5	*Node configuration* Each node shall be equipped with the following components: • Two (2) state-of-the-art, server-grade CPUs. The node must provide a minimum of thirteen (32) physical CPU cores for every GPU installed within the node. • Four (4) or Eight (8) high-end GPUs designed for high-performance computing and AI workloads. • Internal GPU fabric: The GPUs shall be interconnected inside the node through a high-bandwidth, low-latency interface that enables tight coupling and efficient peer-to-peer communication, supporting unified memory access and workload sharing across all GPUs' node.	MRQ
5.3.2-6	CPU Technology The processor architecture implemented within the proposed system shall be based on either x86_64 or ARM 64-bit (AArch64) instruction set architecture, ensuring full compatibility with standard AI and HPC software toolchains. Only general-purpose server-grade CPUs shall be accepted, supporting hardware-assisted virtualization (e.g., Intel VT-x, AMD-V, ARM Virtualization Extensions), IOMMU for secure device assignment, containerization (e.g., KVM), and confidential computing capabilities (e.g., Intel TDX, AMD SEV, ARM Realm Management Extension) for secure execution of sensitive workloads.	MRQ

5.3.2-7	*GPU technology* The proposed system shall integrate state-of-the-art GPU accelerators explicitly designed to support AI training and high-performance scientific simulation. Each GPU shall meet the following minimum technical specifications: • Support hardware-level optimization for AI training, including native capabilities for tensor operations and mixed-precision arithmetic (e.g., FP16, BF16, FP8, FP4, etc.), with particular applicability to training Large Language Models (LLMs). • Support hardware-level optimization for scientific computing, including native capabilities for double-precision floating-point arithmetic (FP64), with particular relevance to HPC workloads, numerical simulations, and advanced computational models. • Be equipped with a minimum of 140 GB of High Bandwidth Memory (HBM) per GPU, ensuring adequate memory bandwidth and capacity to process large-scale deep learning models efficiently. • Support high-bandwidth, low-latency scale-up interconnect among GPUs, enabling peer-to-peer memory access and accelerated collective communication operations. This functionality shall enable efficient execution of large-scale distributed training workloads requiring tight coupling and high throughput among multiple GPU devices. • Only integrated module-based GPU technologies shall be accepted (e.g., NVIDIA SXM, OAM form factor, or equivalent).	MRQ
5.3.2-8	*System memory configuration* Each node shall be provisioned with a minimum of 256 GB of system memory per installed GPU. Memory shall be implemented using DDR5, LP-DDR5, or newer technologies, and configured to maximize effective memory bandwidth, channel utilization, and overall system reliability in order to guarantee sustained performance for memory-intensive computational workloads.	MRQ
5.3.2-9	*Network configuration* Each node shall be equipped with the following network interfaces to ensure high-performance connectivity, system management, and network function offloading capabilities: HPAIF Connectivity: • One (1) or more high-speed NICs, delivering a minimum bandwidth of 400 Gb/s per installed GPU, resulting in an aggregated node-level throughput of at least 1.6 Tb/s or 3.2 Tb/s, respectively for 4-way or 8-way configuration. These NICs shall interface directly with the HPAIF and shall support high-throughput data exchange, ensuring optimal performance for large-scale AI training workloads and tightly coupled GPU communication patterns. CEF Connectivity: • One (1) or more Ethernet NICs, delivering a minimum of 100 Gb/s per installed GPU, resulting in an aggregated node-level throughput of at least	MRQ

	400 Gb/s or 800 Gb/s, respectively for the 4-way or 8-way configuration. The network ports shall be connected to at least two distinct top-of-rack (or leaf switches) to ensure path redundancy and fault tolerance. These NICs shall support the following hardware capabilities: o VLAN offloading in hardware. o VXLAN encapsulation and decapsulation offloading in hardware. o RDMA over Converged Ethernet (RoCE) versions 1 and 2. o RoCE support over overlay networks (VXLAN/NVGRE). o NVMe over TCP (NVMe/TCP) target offload capabilities. o TCP/UDP/IP stateless offloading. o Single Root I/O Virtualization (SR-IOV). IB Management Connectivity: • One (1) NIC featuring one (1) Ethernet port operating at 10/25 Gb/s. The NIC shall support remote boot over Ethernet leveraging industry-standard mechanisms such as PXE, HTTP Boot (UEFI-based), NVMe/TCP Boot. OOB Management Connectivity: • One (1) NIC featuring a single (1) Ethernet port operating at 1 Gb/s (Base-T) providing access to the OOB Management Network via the BMC. Note: The OOB and IB management interfaces may share the same physical port, provided that all port connectivity and functional requirements are fully met.	
5.3.2-10	*Local storage* Each node shall be equipped with the following local storage components: • One (1) SSD drive with a minimum of 200 GB of usable capacity, dedicated to the operating system and critical system services. • One (1) or more SSD drives with a minimum aggregated usable capacity of 5 TB. These drives shall be locally direct attached and shall deliver sustained performance suitable for AI checkpoint and intensive I/O workloads.	MRQ

5.4 Conventional Compute Partition

The Conventional Compute Partition is composed of Compute Nodes equipped exclusively with general-purpose CPUs. This partition is designed to support a wide range of workloads that do not require hardware acceleration, including serial and moderately parallel applications, pre/post-processing tasks, and orchestration services. The infrastructure must ensure reliable performance, efficient resource utilization, and seamless integration with the broader HPC and cloud ecosystem.

5.4.1 Cloud Application Sub-Partition

The Cloud Application sub-partition is optimized for hosting cloud-native services and containerized applications. These nodes must support dynamic resource allocation, high availability, and integration with orchestration platforms. The infrastructure should enable flexible deployment of microservices, APIs, and cloud applications, ensuring scalability, resilience, and compatibility with modern cloud architectures. The following requirements outline the capabilities expected to support cloud application workloads effectively.

Req.	Description	Category
5.4.1-1	*General requirements* The Cloud Application Nodes shall collectively provide of twelve thousand (12,000) physical cores.	MRQ
5.4.1-2	*Common requirements* The nodes shall implement the requirements provided in Req. 5.1.1-3.	MRQ
5.4.1-3	*Cooling technology* The Cloud Application Nodes shall employ Direct Liquid Cooling (DLC) solutions capable of dissipating at least 65% of the system's total TDP through the liquid cooling loop. The DLC implementation shall be designed to ensure thermal stability under sustained high-load operation, minimize reliance on air cooling, and enable efficient heat recovery strategies.	MRQ
5.4.1-4	*Target cooling technology* The Cloud Application Nodes should employ Direct Liquid Cooling (DLC) solutions capable of dissipating at least 90% of the system's total TDP through the liquid cooling loop. The DLC implementation should be designed to ensure thermal stability under sustained high-load operation, minimize reliance on air cooling, and enable efficient heat recovery strategies.	TRQ
5.4.1-5	*CPU configuration* Each node shall be equipped with two (2) server-grade CPUs based on the x86_64 architecture. The processors shall feature a high number of physical cores to ensure high computational throughput and energy-efficient performance for cloud-native workloads. Each CPU shall support hardware-assisted virtualization and containerization technologies (e.g., Intel VT-x/AMD-V, VT-d/IOMMU), as well as confidential computing capabilities (e.g., Intel TDX, AMD SEV) to enable secure execution of sensitive workloads in isolated environments.	MRQ
5.4.1-6	*System memory configuration* Each node shall be equipped with a minimum of 2.5 GB of system memory for each physical core in the node. Memory shall be implemented using DDR5, LP-DDR5, or newer technologies, and configured to maximize effective memory bandwidth, channel utilization, and overall system reliability in order to guarantee sustained performance for memory-intensive computational workloads.	MRQ
5.4.1-7	*Network requirements* Each node shall be equipped with the following network interfaces to ensure high-performance connectivity, system management, and network function offloading capabilities:	MRQ

	CEF Connectivity: • One (1) NIC with two (2) Ethernet ports, each delivering a minimum of 100 Gb/s, for an aggregated throughput of 200 Gb/s. The two network ports shall be connected to distinct top-of-rack or leaf switches to ensure path redundancy and fault tolerance. The NIC shall support the following hardware capabilities: o VLAN offloading in hardware. o VXLAN encapsulation and decapsulation offloading in hardware. o RoCE versions 1 and 2. o RoCE support over overlay networks (VXLAN/NVGRE). o NVMe/TCP target offload capabilities. o TCP/UDP/IP stateless offloading. o SR-IOV. IB Management Connectivity: • One (1) NIC featuring one (1) Ethernet port operating at 10/25 Gb/s. The NIC shall support remote boot over Ethernet leveraging industry-standard mechanisms such as PXE, HTTP Boot (UEFI-based), NVMe/TCP Boot. OOB Management Connectivity: • One (1) NIC featuring a single (1) Ethernet port operating at 1 Gb/s (Base-T) providing access to the OOB Management Network via the BMC. Note: The OOB and IB management interfaces may share the same physical port, provided that all port connectivity and functional requirements are fully met.	
5.4.1-8	*Local storage* Each node shall be equipped with a RAID 1 (mirroring) configuration using SSD drives, ensuring redundancy and providing a minimum of 200 GB of usable capacity, dedicated to the operating system and critical system services.	MRQ

5.4.2 Industrial HPC Simulation Sub-Partition

The Industrial HPC Simulation Partition is specifically tailored for executing high-performance industrial computing simulations. These nodes must be optimized for tightly coupled parallel workloads, numerical modeling, and scientific computing tasks that rely on deterministic performance and efficient inter-node communication. The infrastructure must support scalable execution of MPI-based applications and ensure consistent throughput across simulation environments. The partition is internally interconnected via a high-bandwidth, low-latency fabric (HPCSF), enabling fast and reliable data exchange essential for distributed HPC workloads.

Req.	**Description**	**Category**
5.4.2-1	*General requirements* The Industrial HPC Simulation Nodes must include Forty (40) physical nodes.	MRQ
5.4.2-2	*Common requirements* The nodes shall implement the requirements provided in Req. 5.1.1-3.	MRQ

5.4.2-3	*Cooling technology* The Industrial HPC Simulation Nodes shall employ Direct Liquid Cooling (DLC) solutions capable of dissipating at least 65% of the system's total TDP through the liquid cooling loop. The DLC implementation shall be designed to ensure thermal stability under sustained high-load operation, minimize reliance on air cooling, and enable efficient heat recovery strategies.	MRQ
5.4.2-4	*Target cooling technology* The Industrial HPC Simulation Nodes should employ Direct Liquid Cooling (DLC) solutions capable of dissipating at least 90% of the system's total TDP through the liquid cooling loop. The DLC implementation should be designed to ensure thermal stability under sustained high-load operation, minimize reliance on air cooling, and enable efficient heat recovery strategies.	TRQ
5.4.2-5	*CPU configuration* Each node shall be equipped with two (2) server-grade CPUs based on the x86_64 architecture. The processors shall feature cores with high single-thread performance to ensure optimal execution of industrial workloads that require low-latency response and benefit from fast sequential processing. Each CPU shall support hardware-assisted virtualization and containerization technologies (e.g., Intel VT-x/AMD-V, VT-d/IOMMU), as well as confidential computing capabilities (e.g., Intel TDX, AMD SEV) to enable secure execution of sensitive workloads in isolated environments.	MRQ
6.4.2-6	*System memory configuration* Each node shall be equipped with a minimum of 768 GB of system memory. Memory shall be implemented using DDR5, LP-DDR5, or newer technologies, and configured to maximize effective memory bandwidth, channel utilization, and overall system reliability in order to guarantee sustained performance for memory-intensive computational workloads.	MRQ
5.4.2-7	*Network requirements* Each node shall be equipped with the following network interfaces to ensure high-performance connectivity, system management, and network function offloading capabilities: HPCSF Connectivity: • One (1) host-channel adapters with one (1) InfiniBand port, delivering a minimum bandwidth of 400 Gb/s. This NIC shall interface directly with the HPCSF and shall support low-latency data exchange, ensuring optimal performance for HPC workloads. CEF Connectivity: • One (1) NIC with two (2) Ethernet ports, each delivering a minimum of 100 Gb/s, for an aggregated throughput of 200 Gb/s. The two network ports shall be connected to distinct top-of-rack or leaf switches to ensure path	MRQ

	redundancy and fault tolerance. The NIC shall support the following hardware capabilities: o VLAN offloading in hardware. o VXLAN encapsulation and decapsulation offloading in hardware. o RoCE versions 1 and 2. o RoCE support over overlay networks (VXLAN/NVGRE). o NVMe/TCP target offload capabilities. o TCP/UDP/IP stateless offloading. o SR-IOV. IB Management Connectivity: • One (1) NIC featuring one (1) Ethernet port operating at 10/25 Gb/s. The NIC shall support remote boot over Ethernet leveraging industry-standard mechanisms such as PXE, HTTP Boot (UEFI-based), NVMe/TCP Boot. OOB Management Connectivity: • One (1) NIC featuring a single (1) Ethernet port operating at 1 Gb/s (Base-T) providing access to the OOB Management Network via the BMC. Note: The OOB and IB management interfaces may share the same physical port, provided that all port connectivity and functional requirements are fully met.	
5.4.2-8	*Local storage* Each node shall be equipped with a RAID 1 (mirroring) configuration using SSD drives, ensuring redundancy and providing a minimum of 200 GB of usable capacity, dedicated to the operating system and critical system services.	MRQ

5.5 Management partition

The Management Partition shall be structured into three functionally distinct sub-partitions: the Master Partition, the Service Partition (Cloud Controller & Container Orchestrator Nodes), and the Management Storage subsystem. Each of these components serves a critical role in ensuring the reliable, secure, and efficient management operation of the overall infrastructure.

5.5.1 Master partition

The Master Partition shall comprise a logically and physically distinct group of nodes responsible for end-to-end cluster management operations. These nodes shall serve as the main interface to CINECA's system administrators, supporting secure interaction with core infrastructure services such as system monitoring, configuration management, centralized authentication, and logging subsystems. The design shall ensure high availability, operational resilience, and secure administrative access.

Req.	**Description**	**Category**
5.5.1-1	*General requirements* The Master partition shall include at minimum two (2) physical nodes.	MRQ

5.5.1-2	*Connection with CINECA backbone* The Master Nodes shall be connected to the CINECA backbone of CINECA, which is based on NVIDIA Spectrum SN4700 named as SWT3 and SWT4 as shown in Figure 8. The Supplier shall provide and install all required physical interconnect components, including but not limited to high-speed optical transceivers (of type 100GBase-DR1), compatible cabling (e.g., single-mode optical fibre), and passive infrastructure elements needed to complete the physical connection between the Master Nodes and the backbone switches (SWT3 & SWT4). The network architecture shall ensure both performance optimization and resilience, connecting the Master Nodes to distinct switches to support path redundancy and fault tolerance. Integration and logical configuration of the interfaces to the CINECA backbone shall be performed in close coordination with CINECA's technical personnel, ensuring alignment with the existing Layer 2/3 routing, VLAN segmentation, and access control policies. The Master Nodes shall be configured to participate in the management traffic flows of the entire system infrastructure, acting as the central point for administrative operations, orchestration, and telemetry collection.	MRQ
5.5.1-3	*Common requirements* The nodes shall implement the requirements provided in Req. 5.1.1-3.	MRQ
5.5.1-4	*CPU configuration* Each node shall be equipped one (1) or two (2) server-grade CPUs based on the x86_64 architecture. The processors shall provide sufficient computational capabilities to support system management tasks and shall include full support for hardware-assisted virtualization and containerization technologies (e.g., Intel VT-x/AMD-V, VT-d/IOMMU), as well as confidential computing capabilities (e.g., Intel TDX, AMD SEV) to enable secure execution of sensitive workloads in isolated environments.	MRQ
5.5.1-5	*System memory configuration* Each node shall be equipped with a minimum of 384 GB of system memory. This memory capacity shall ensure adequate resources to support system management functions, critical infrastructure services, and the execution of virtualized or containerized workloads. Memory shall be implemented using DDR5, LP-DDR5, or newer technologies, and configured to maximize effective memory bandwidth, channel utilization, and overall system reliability.	MRQ
5.5.1-6	*Network requirements* Each node shall be equipped with the following network interfaces to ensure high-performance connectivity, system management, and network function offloading capabilities and shown in Figure 8: HPCSF Connectivity:	MRQ

	• One (1) HCA with one (1) InfiniBand port providing a minimum bandwidth of 100 Gb/s, dedicated to interfacing directly with the HPCSF for debugging, monitoring, and low-level diagnostics. This interface shall allow non-intrusive access to CPU communication paths and fabric-level telemetry, facilitating system-level troubleshooting and performance analysis without interfering with primary AI workload traffic. HPAIF & CEF Connectivity: • One (1) NIC with two (2) Ethernet port providing a minimum bandwidth of 100 Gb/s for each fabric, dedicated to interfacing directly with the HPAIF and CEF for debugging, monitoring, and low-level diagnostics. This interface shall allow non-intrusive access to GPU and storage communication paths and fabric-level telemetry, facilitating system-level troubleshooting and performance analysis without interfering with primary AI and storage workload traffic. In case the required ports for HPAIF are not sufficient and the fabric consists of a single switch; it is sufficient to connect the management port of the switch to the management network. CINECA's admin Connectivity: • One (1) NIC with two (2) Ethernet ports, each delivering a minimum of 100 Gb/s. The NIC shall interface with CINECA's admin network through the CINECA backbone switches. For this reason, it shall be equipped with optical transceivers of type 100GBase-DR1, operating over single-mode fibre (SMF). The two network ports shall be connected to distinct CINECA backbone switches to ensure path redundancy and fault tolerance. IB Management Connectivity: • One (1) NIC with two (2) Ethernet ports, each delivering a minimum of 100 Gb/s. The NIC shall be dedicated to connectivity with the IB Management Network of the system. The two network ports shall be connected to distinct top-of-rack or leaf switches to ensure path redundancy and fault tolerance. These NICs shall support the following hardware capabilities: o VLAN offloading in hardware. o VXLAN encapsulation and decapsulation offloading in hardware. o RoCE versions 1 and 2. o RoCE support over overlay networks (VXLAN/NVGRE). o NVMe/TCP target offload capabilities. o TCP/UDP/IP stateless offloading. o SR-IOV. CINECA's OOB Connectivity: • One (1) NIC featuring one (1) Ethernet port operating at 1 Gb/s (Base-T) providing access to the OOB Management Network via the BMC. Note: The OOB and IB management interfaces may share the same physical port, provided that all port connectivity and functional requirements are fully met.	
5.5.1-7	*Local storage*	MRQ

	Each node shall be equipped with a RAID 1 (mirroring) configuration using SSD drives, ensuring redundancy and providing a minimum of 1 TB of usable capacity, dedicated to the operating system and critical system services.	

5.5.2 Service partition

The Service Partition is composed of Cloud Controllers and Container Orchestrators as shown in Figure 6. It shall consist of a dedicated set of nodes allocated to host system-level services, these include, but are not limited to, Cloud Controllers, Container Orchestrators, system monitoring frameworks, and auxiliary infrastructure services. These nodes shall provide a robust and highly available runtime environment with performance isolation and scalability for critical service workloads. The Service Partition shall also facilitate seamless integration with external infrastructure components such as identity and access management systems, license servers, and software repositories. This integration ensures centralized governance, secure operations, and uniform service delivery across the platform.

Req.	**Description**	**Category**
5.5.2-1	*General requirements* The Service partition shall include no fewer than six (6) physical nodes in cluster configuration. To optimize resource utilization, candidates are encouraged to adopt virtualization technologies, such as virtual machines or container-based orchestrator, to consolidate services where appropriate, provided that the functional and performance requirements are fully met, and the minimum node count is respected.	MRQ
5.5.2-3	*Common requirements* The nodes shall implement the requirements provided in Req. 5.1.1-3.	MRQ
5.5.2-4	*CPU configuration* Each node shall be equipped one (1) or two (2) server-grade CPUs based on the x86_64 architecture. The processors shall provide sufficient computational capabilities to support system management tasks and shall include full support for hardware-assisted virtualization and containerization technologies (e.g., Intel VT-x/AMD-V, VT-d/IOMMU), as well as confidential computing capabilities (e.g., Intel TDX, AMD SEV) to enable secure execution of sensitive workloads in isolated environments.	MRQ
5.5.2-5	*System memory configuration* Each node shall be equipped with a minimum of 384 GB of system memory. This memory capacity shall ensure adequate resources to support system management functions, critical infrastructure services, and the execution of virtualized or containerized workloads. Memory shall be implemented using DDR5, LP-DDR5, or	MRQ

	newer technologies, and configured to maximize effective memory bandwidth, channel utilization, and overall system reliability.	
5.5.2-6	*Network requirements* Each node shall be equipped with the following network interfaces to ensure high-performance connectivity, system management, and network function offloading capabilities: CEF Connectivity: • One (1) NIC with two (2) Ethernet ports each delivering a minimum of 100 Gb/s. The two network ports shall be connected to distinct top-of-rack or leaf switches to ensure path redundancy and fault tolerance. These NICs shall support the following hardware capabilities: o VLAN offloading in hardware. o VXLAN encapsulation and decapsulation offloading in hardware. o RoCE versions 1 and 2. o RoCE support over overlay networks (VXLAN/NVGRE). o NVMe/TCP target offload capabilities. o TCP/UDP/IP stateless offloading. o SR-IOV. IB Management Connectivity: • One (1) NIC with two (2) Ethernet ports, each delivering a minimum of 100 Gb/s. The NIC shall be dedicated to connectivity with the IB Management Network of the system. The two network ports shall be connected to distinct top-of-rack or leaf switches to ensure path redundancy and fault tolerance. These NICs shall support the following hardware capabilities: o VLAN offloading in hardware. o VXLAN encapsulation and decapsulation offloading in hardware. o RoCE versions 1 and 2. o RoCE support over overlay networks (VXLAN/NVGRE). o NVMe/TCP target offload capabilities. o TCP/UDP/IP stateless offloading. o SR-IOV. OOB Management Connectivity: • One (1) NIC featuring one (1) Ethernet port operating at 1 Gb/s (Base-T) providing access to the OOB Management Network via the BMC. Note: The OOB and IB management interfaces may share the same physical port, provided that all port connectivity and functional requirements are fully met.	MRQ
5.5.2-7	*Local storage* Each node shall be equipped with a RAID 1 (mirroring) configuration using SSD drives, ensuring redundancy and providing a minimum of 800 GB of usable capacity, dedicated to the operating system and critical system services.	MRQ
5.5.2-8	*Performance*	MRQ

	The number and configuration of Service Nodes shall be dimensioned to ensure compliance with all required performance levels for system installation, reboot, and provisioning operations. The infrastructure shall also guarantee the efficient collection, storage, and processing of all system metrics and logs. Service Nodes shall deliver consistently high-performance during data query operations and shall support, without degradation, all activities related to system management, troubleshooting, accounting, and security assessments. The solution shall be designed to maintain optimal responsiveness and reliability under full system load and concurrent access scenarios.	
5.5.2-9	*High Availability* The Service Partition shall include all necessary hardware and software components to deploy critical system services in a high availability configuration. Service Nodes shall operate in cluster mode, ensuring both hardware and software-level fault tolerance. The architecture shall guarantee uninterrupted operation of all essential services in the event of a single-node or dual-node failure, maintaining appropriate levels of service continuity and operational efficiency. The Service Partition may be organized using virtualization and/or containerization technologies. When adopted, solutions such as Proxmox VE (or equivalent) for virtualization and Kubernetes (or equivalent) for container orchestration can be implemented to ensure flexibility, scalability, and efficient cluster management. These technologies enable simplified lifecycle operations, resource optimization, and rapid recovery in case of failures. For software components that are not certified for virtualization or containerization, the architecture shall provide dedicated node pairs configured in high availability mode to guarantee service continuity and fault tolerance.	MRQ

5.5.3 Management storage

The Management Storage subsystem shall provide a shared, dedicated storage infrastructure exclusively accessible by the Management Nodes. It shall be designed for the reliable collection, storage, and management of system-level metadata, configuration files, telemetry streams, and operational logs. This component shall support centralized, secure, and policy-compliant data retention practices, enabling effective system diagnostics, auditability, performance monitoring, and disaster recovery. The storage layer shall implement fine-grained access control mechanisms and be tightly integrated with the system's monitoring, provisioning, and configuration management toolchain to support automated lifecycle operations and enforce operational policies.

Req.	**Description**	**Category**
5.5.3-1	*General requirements* The Management storage shall fulfil the following functional requirements: • Host all management software components and related data required for system operation and orchestration.	MRQ

	• Store all management-related databases, including a daily differential backup history retained for a minimum of one (1) year. • Retain aggregated system logs collected from all node partitions for a minimum of one (1) year. • Retain aggregated audit logs from Compute, Front-End, and Management Nodes for a minimum of two (2) years. • Store all performance and functional metrics collected from all system nodes and equipment for a minimum of two (2) years.	
5.5.3-2	*Capacity* The Management storage shall provide a minimum usable capacity of 400 TB, exclusively dedicated to supporting system management operations and monitoring data.	MRQ
5.5.3-3	*Network requirements* The Management storage shall be connected to the Management Network and accessible by all Management Nodes as shown in Figure 8. It shall be uniformly shareable across all Management Nodes to ensure continuous availability of services. The Management storage shall be physically and logically separated from the storage subsystems assigned to Compute, Front-End, and other system partitions.	MRQ
5.5.3-4	*High Availability and Fault Tolerance* The storage solution shall be designed for high availability and be resilient to the simultaneous failure of at least two independent basic hardware blocks (e.g., storage nodes, controllers, or disk enclosures), without causing service disruption or data loss.	MRQ
5.5.3-5	*Integration with Service Nodes* In configurations where Service Nodes are clustered using Kubernetes or a KVM-based virtualization platform (e.g., Proxmox VE), the Management storage shall be implemented as a distributed, software-defined storage system.	MRQ

5.6 Storage infrastructure

5.6.1 Data Lake

The Data Lake is a scalable, high-performance storage subsystem designed to support the needs of AI-driven and HPC workloads. It is implemented as a full-flash architecture to ensure low-latency, high-throughput, and reliable performance across all operational scenarios. The platform provides large-scale capacity, efficient data store, and seamless access from all system partitions. It supports a wide range of protocols and interfaces including file and parallel access while enabling secure multi tenancy, centralized authentication, and policy-based resource control. All the required storage features shall be implemented and confirmed as operational by the completion of the acceptance phase.

Req.	Description	Category

5.6.1-1	*Technology* The infrastructure shall be based on a full-flash architecture, ensuring low latency, high throughput, and consistent performance across all workloads.	MRQ
5.6.1-2	*Capacity* The system shall provide a minimum of 7 PB of usable capacity (net space), available for production use. The Data Lake can achieve the required usable capacity through data reduction mechanisms such as inline compression, deduplication, or similar technologies. However, the Candidate shall: • Explicitly declare the level of usable capacity (net space) being committed, specifying whether this capacity is based on data reduction or raw allocation. • Guarantee that, in the event the declared net capacity is not achieved during operational use, additional capacity will be provided at no cost to meet the committed usable space requirements. Data reduction must be performed inline, transparently by the system prior to data being written to drives and must not require manual configuration or ongoing management by system administrators. All throughput and IOPS declared in the proposal must be presented inclusive of any performance introduced by data reduction mechanisms, ensuring accurate representation of real-world performance.	MRQ
5.6.1-3	*Performance* The architecture shall guarantee a minimum sustained end-to-end throughput of: • *Read throughput*: at least 200 GB/s sustained under normal operating conditions. • *Peak write throughput*: at least 60 GB/s sustained for checkpoint operations of no less than 15 minutes. • *Write throughput*: at least 30 GB/s sustained under normal operating conditions.	MRQ
5.6.1-4	*Core Ethernet Fabric connection* The Data Lake shall be interconnected via CEF to all partitions of the System. Each Compute and Management Node shall be capable of mounting and accessing data residing on the Data Lake.	MRQ
5.6.1-5	*System architecture* • The Data Lake shall support client access over IPv4 protocol to ensure compatibility with IP network environments. • The Data Lake shall employ a scale-out architecture to enable seamless expansion and to support growing workload and data requirements. • The Data Lake shall allow scaling of capacity and performance resources, enabling flexible adaptation to evolving operational demands.	MRQ

	• The Data Lake shall provide transparent failover capabilities in the event of a front-end node or network failure, ensuring uninterrupted data access and service continuity. • The namespaces of the Data Lake shall be scalable to support up to 100 PB.	
5.6.1-6	*Accessibility & multi-tenancy* • All namespaces shall be accessible from every partition within the System, ensuring unified data visibility and interoperability. • The namespaces shall be exposed to client nodes through a POSIX-compliant interface to ensure compatibility with standard file system operations. • The Data Lake shall support integration with LDAP services for centralized authentication and user management. • The offered Data Lake shall support the capability to logically partition storage resources into isolated tenants, each with separate data paths, to serve multiple clients securely and independently. This multi-tenancy design shall ensure data isolation, access control enforcement, and resource allocation per tenant, enabling secure and scalable service delivery across different organizational entities or projects.	MRQ
5.6.1-7	*High availability & resiliency* • The Data Lake shall provide full redundancy, eliminating any single point of failure to ensure continuous availability and operational resilience. • The Data Lake shall implement transparent failover mechanisms, enabling uninterrupted service in the event of a node or network failure. • The data rebuild process shall support fail-in-place functionality and initiate automatically upon fault detection, without requiring hardware replacement or manual operator intervention. • All software and firmware updates shall be performed in a non-disruptive manner, ensuring uninterrupted service availability to all users during maintenance operations. • In the event of a component failure within the Data Lake, the proposed System shall execute an automated fault recovery procedure that maintains data integrity and ensures continued access to all stored information without operational disruption.	MRQ
5.6.1-8	*Parallel filesystem* • The Data Lake shall support a parallel file system, or equivalent technology, to enable concurrent read and write operations on the same file across multiple data front-end nodes. This capability shall facilitate high-performance data access through coordinated, simultaneous I/O operations between client nodes and the Data Lake. • The client shall support multipath connectivity across multiple network links to ensure resilient and high-throughput access.	MRQ
5.6.1-9	*NFS support*	TRQ

	• The Data Lake should support the NFS protocols version 3 and version 4.1 to ensure compatibility with a broad range of client systems. • The Data Lake should support both POSIX Access Control Lists (ACLs) and NFS v4.1 ACLs to enforce fine-grained permissions on shared file systems. • The Data Lake should support NFS over Remote Direct Memory Access (RDMA) for NFS v4.1, enabling low-latency, high-throughput data access.	
5.6.1-10	*S3 support* • The Data Lake shall support a maximum object size of 4 TB or greater, ensuring accommodation of large data objects. • The solution shall support more than 100 buckets per namespace, enabling scalable namespace management. • The Data Lake shall implement S3-compatible identity and bucket policies to enforce access control and security at the bucket level. • The solution shall support multipart uploads, allowing efficient and reliable upload of large objects by dividing them into parts. • The Data Lake shall provide Object Lock functionality to enable write-once-read-many (WORM) protection for compliance and data retention requirements.	MRQ
5.6.1-11	*Multi-protocol support* • The Data Lake should support object access via S3-compatible APIs, ensuring interoperability with standard S3 clients and applications. • The Data Lake should support file access via NFS protocol. • The Data Lake should support client access over IPv4 network protocols to ensure compatibility with existing network infrastructure. • The Data Lake should enforce configurable quotas to regulate and limit resource consumption according to defined policies. • Files created and accessed via parallel file system and NFS protocol should be accessible and represented as objects via the S3 interface, ensuring seamless data consistency between protocols. • Objects created and accessed via the S3 interface should be accessible as files via the parallel file system and NFS protocol, enabling bi-directional interoperability. • The Data Lake should support asynchronous replication of data between geographically distributed sites, utilizing the same underlying technology to ensure data availability and disaster recovery.	TRQ
5.6.1-12	*Security features* • The Data Lake shall provide the capability to encrypt all data prior to being written to persistent storage media, ensuring confidentiality at rest. • The Data Lake shall support encryption in flight leveraging industry-standard mechanisms to protect data during transmission. • The Data Lake shall support the creation of immutable snapshots to ensure data integrity and support protection against Ransomware threats, with the capability to prevent deletion or modification for a defined retention period. • The Data Lake shall provide the capability to clone snapshots and make the cloned instances available in read/write mode, enabling rapid data	MRQ

	reuse for testing, development, or recovery purposes without impacting the original data set.	
5.6.1-13	*Management & audit* • The Data Lake shall be integrated with the Management Networks to facilitate administrative operations and configuration activities. • The management partition of the System shall be granted access to the management interfaces of all storage infrastructures to ensure centralized control and monitoring capabilities. • The Data Lake shall support management through multiple interfaces such as Command Line Interface), Graphical User Interface over HTTPS, or RESTful API, to ensure operational flexibility and automation capabilities. • The Data Lake shall provide a centralized monitoring and analytics service capable of aggregating health status, performance metrics, and operational data across all components of the Data Lake. • The monitoring and analytics services shall include functionality for proactive detection of anomalies and operational issues, enabling early intervention and reduced downtime. • The Data Lake shall support secure collection, transmission, and storage of log and analytics data, ensuring compliance with relevant data protection and security standards. • The Data Lake shall implement RBAC to manage and restrict access to both data and administrative functions based on user roles and permissions. • The Data Lake shall integrate with LDAP to authenticate administrative users via an external directory service. • The Data Lake shall generate and maintain comprehensive audit logs of all administrative access events, including user identity, timestamp, and actions performed. • The Data Lake shall generate and maintain detailed audit logs of client operations, enabling traceability, security auditing, and compliance verification.	MRQ
5.6.1-14	*File system metadata* The offered Data Lake shall support the capability to expose file system metadata in a structured, queryable format, making it accessible as database tables to enable advanced search, reporting, and integration with analytics workflows.	MRQ
5.6.1-15	*Quality of Services (QoS)* The Data Lake shall support QoS mechanisms to throttle and control performance metrics, including IOPS and throughput.	MRQ
5.6.1-16	*Advanced analytics* • The Data Lake should natively support the storage of metrics, data, and metadata within an integrated columnar database, fully embedded and managed within the platform architecture.	TRQ

	• The integrated database should provide seamless interoperability with modern Big Data processing frameworks such as Apache Spark, Trino, Dremio, etc. • The database should support ACID (Atomicity, Consistency, Isolation, Durability) properties and be engineered to handle several transactions per second and scaling to support high query throughput. • The system should include a built-in tabular database capable of supporting schema evolution, allowing flexible extension to accommodate data types such as images and video, alongside structured data related to files and objects. • The Data Lake should integrate an event broker framework, such as Apache Kafka, capable of capturing and emitting file system events and metadata updates. This integration should enable the configuration of event-driven triggers to automate data processing workflows, notifications, and downstream application integrations.	
5.6.1-17	*Software licensing and support* • The software licenses and support services, including the replacement of all hardware components regardless of flash memory wear, shall be provided for the Data Lake. • Software licenses shall be transferable to replacement hardware, including hardware of subsequent generations or upgraded models, without additional licensing costs. • The offer shall include software licenses for backup and recovery services as an integral component of the proposed Data Lake solution.	MRQ

5.6.2 System Storage

This section describes the Ceph-based storage subsystem supporting the OpenStack and Orchestrated computing partition environments. The infrastructure shall provide persistent and resilient storage services for virtual machines and ancillary components, including container images and orchestration metadata. The architecture must include a sufficient number of Ceph OSD nodes for data storage and MON nodes for cluster coordination and health monitoring, ensuring scalability, fault tolerance, and consistent performance across all cloud-native operations.

Req.	**Description**	**Category**
5.6.2-1	*Common requirements* The nodes shall implement the requirements provided in Req. 5.1.1-3. These nodes will serve as part of the Ceph storage infrastructure, supporting the deployment of essential components such as OSD (Object Storage Daemon) and MON (Monitor) services. Their configuration shall ensure reliable data placement, cluster health monitoring, and scalable integration with OpenStack and Orchestrated computing partition environments.	MRQ
5.6.2-2	*CPU configuration*	MRQ

	Each node shall be equipped one (1) or two (2) server-grade CPUs based on the x86_64 architecture.	
5.6.2-3	*System memory configuration* Each node shall be equipped with a minimum of 384 GB of system memory. This memory capacity shall ensure adequate resources to support the storage infrastructure. Memory shall be implemented using DDR5, LP-DDR5, or newer technologies, and configured to maximize effective memory bandwidth, channel utilization, and overall system reliability.	MRQ
5.6.2-4	*Network requirements* Each node shall be equipped with the following network interfaces to ensure high-performance connectivity, system management, and network function offloading capabilities: CEF Connectivity: • One (1) NIC with two (2) Ethernet ports, each delivering a minimum of 100 Gb/s, for an aggregated throughput of 200 Gb/s. The two network ports shall be connected to distinct top-of-rack or leaf switches to ensure path redundancy and fault tolerance. These NICs shall support the following hardware capabilities: o VLAN offloading in hardware. o VXLAN encapsulation and decapsulation offloading in hardware. o RDMA over Converged Ethernet (RoCE) versions 1 and 2. o RoCE support over overlay networks (VXLAN/NVGRE). o NVMe over TCP (NVMe/TCP) target offload capabilities. o TCP/UDP/IP stateless offloading. o Single Root I/O Virtualization (SR-IOV). IB Management Connectivity: • One (1) NIC featuring two (2) Ethernet ports operating at 10/25 Gb/s. The two network ports shall be connected to distinct top-of-rack or leaf switches to ensure path redundancy and fault tolerance. The NIC shall support remote boot over Ethernet leveraging industry-standard mechanisms such as PXE (Preboot Execution Environment), HTTP Boot (UEFI-based), and NVMe over TCP (NVMe/TCP) Boot. OOB Management Connectivity: • One (1) NIC featuring a single (1) Ethernet port operating at 1 Gb/s (Base-T) providing access to the OOB Management Network via the BMC. Note: The OOB and IB management interfaces may share the same physical port, provided that all port connectivity and functional requirements are fully met.	MRQ
5.6.2-5	*Ceph OSD Nodes* The solution must provide a number of nodes adequate to the offered Ceph storage.	MRQ

	Those specific nodes must also provide a capacity optimized OSD nodes for a total capacity of 600 TB net space available. The offered Ceph solution will be used as block service for the OpenStack and Kubernates infrastructure. Hardware design and architecture of OSD nodes needs to take this into account. Moreover, these nodes must be equipped with: • x2 SSD drives ≥ 200 GB units for journaling. • Number of cores enough to support the IO frontend and OSDs as per CEPH best practices. • Planar and HBAs bandwidth must be adequate and balanced against the mix of SSD drives, SATA SSD drives, and NL-SATA disks hosted, to optimize the IO flows.	
5.6.2-6	*Ceph MON and MDS Nodes* The solution must provide a number of nodes adequate to the offered Ceph storage. Moreover, these nodes must be equipped with: • x2 SSD drives in RAID1 configuration with net space available ≥ 3 TB.	MRQ

5.7 System software

This section defines the requirements for the deployment, scheduling, orchestration, and software environment of the infrastructure. It covers the integration of OpenStack and Orchestrated computing partition platforms, operating system standards, authentication services, AI frameworks, performance monitoring tools, and development libraries essential for managing and executing AI/HPC workloads efficiently.

Req.	Description	Category
5.7-1	*Resource provisioning platforms* The proposed infrastructure shall support the simultaneous operation of both OpenStack and Orchestrated computing partition environments. All system partitions must be capable of hosting services and workloads managed by either framework, allowing flexible and dynamic allocation of resources. The architecture shall enable coexistence and interoperability between the two platforms, ensuring that virtual machines, containers, and associated services can be provisioned, monitored, and scaled independently or in coordination, according to operational needs and workload characteristics.	MRQ
5.7-2	*OpenStack platform* The proposed solution shall implement OpenStack as the cloud orchestration platform for managing virtualized resources. The deployment must be configured in high-availability mode and support large-scale, multi-tenant environments. The	MRQ

	orchestrator shall enable dynamic provisioning and lifecycle management of compute, storage, and networking services across all partitions. The infrastructure shall include core OpenStack components such as Nova (compute), Neutron (networking), and Cinder (volume management) to support orchestration and resource provisioning. The architecture shall not be limited to these services; it must allow for the deployment of additional OpenStack components as needed to support evolving operational requirements, integration scenarios, or advanced cloud functionalities. The system design shall ensure compatibility and scalability across the full OpenStack ecosystem. The system must support flexible resource allocation, allowing dynamic reconfiguration of node assignments between OpenStack and other partitions in response to workload demands or operational policies.	
5.7-3	*Orchestrated computing partition platform* The proposed solution shall also implement a Kubernetes-based orchestration platform or equivalent for managing both CPU and GPU resources. This environment shall support advanced workload orchestration features, including GPU-aware scheduling, dynamic resource allocation, user and group quota enforcement, multi-node distributed workload execution, user prioritization, and namespace-level isolation. Integration with centralized authentication systems (e.g., LDAP, Keycloak) and support for RBAC is mandatory. The platform shall provide a unified interface, accessible via CLI and/or web-based dashboards (e.g., Kubernetes Dashboard, custom UIs), enabling configuration, job submission, monitoring, and resource usage tracking. Full compatibility with OCI-compliant container runtimes (such as containerd and/or Docker) is required, along with support for vendor-optimized software environments (e.g., NVIDIA NGC, etc.) shall be ensured. Only solutions natively based on Kubernetes and aligned with the aforementioned specifications, will be considered compliant.	MRQ
5.7-4	*Operating System* • All nodes within the solution shall operate on a supported, enterprise-grade Linux distribution, specifically Ubuntu Pro version 24.04 LTS or higher, or Red Hat Enterprise Linux (RHEL) version 9 or higher, with a 64-bit kernel version 5.14 or newer. The operating system shall include native capabilities for remote system management, support for network boot (e.g., PXE or equivalent protocol), and mechanisms for automated system image delivery and provisioning. • The selected operating system shall be uniformly deployed across all nodes, ensuring full consistency of the software environment and compatibility across the system. • The operating system shall include and support integration with tools for automated configuration management, system monitoring, security baseline enforcement, and patch lifecycle management, leveraging vendor-supported mechanisms such as Canonical Livepatch, Red Hat Satellite, Landscape, or third-party orchestration tools (e.g., Ansible, Puppet). • The operating system must support enterprise-grade security features, including kernel-level access controls, SELinux or AppArmor, FIPS 140-2	MRQ

	compliant cryptographic modules, and compliance with CIS or equivalent hardening benchmarks.	
5.7-5	*Authentication and User Directory Integration* The architecture of the procured infrastructure shall support seamless integration with CINECA's LDAP-based directory service for centralized user authentication and management. This includes native LDAP integration as well as integration through identity and access management platforms such as Keycloak. The integration shall be designed and implemented in a high-availability configuration, ensuring that no single point of failure exists in the authentication and directory access path. Redundancy mechanisms shall be in place to guarantee continuous operation and fault tolerance in the event of directory service disruptions or component failures. The system shall also support federated authentication scenarios, user provisioning, and RBAC as managed through the LDAP directory or via Keycloak.	MRQ
5.7-6	*Container workload support* • The offered solution must support the remote building and execution of containerized applications to effectively support AI workloads. All required components to enable this capability shall be included, such as container runtime software, container image repository storage, and any necessary software licenses. • The system shall provide a mechanism to build and modify containers from the front-end partition, ensuring flexibility and user accessibility for container customization. • The solution shall support at least one standard container format, such as the Open Container Initiative (OCI) Image Specification v1.1.0 or later, or the Singularity Image Format as defined by Sylabs, to ensure interoperability with modern containerized environments. • The solution shall include RBAC to enforce compliance with site-specific security policies and to manage user access to container-related resources and operations.	MRQ
5.7-7	*Support for programming environment standards.* The offered solution shall provide comprehensive support for modern programming environments and AI/ML frameworks, enabling development and execution of advanced machine learning and deep learning workloads. The software stack shall include recent, stable, and optimized versions of the following components: • Python (version 3.10 or newer), with scientific computing and data processing libraries such as NumPy, SciPy, Pandas, scikit-learn, and Matplotlib. • Deep learning frameworks including, but not limited to, PyTorch (version 2.1 or newer), Keras, JAX, ONNX Runtime, and TensorFlow (version 2.13 or newer). • The solution shall provide comprehensive support for GPU and accelerator libraries across multiple hardware vendors, including major GPU architectures. It shall support industry-standard and/or open-source software development kits and libraries for GPU acceleration, such as	MRQ

	CUDA-compatible toolkits and equivalent frameworks enabling optimized deep learning performance and interoperability with popular AI frameworks. • Distributed training frameworks, such as Megatron-LM, Accelerate, Lightning, DeepSpeed, or TorchElastic, supporting scalable multi-GPU and multi-node training. • Containerized environments, with support for vendor-optimized containers (e.g., NVIDIA NGC) and execution engines such as Apptainer/Singularity and integration with container orchestrators for AI workloads. • The solution shall provide a fully integrated accelerator programming stack suitable for heterogeneous environments, including runtime libraries, APIs, compilers, and pre-built container images where applicable. The stack must support GPU offloading, mixed-precision computation (e.g., FP16/BF16, FP8, etc.), and hardware-aware optimizations across multiple hardware vendors. Preference shall be given to open-source libraries and frameworks to ensure transparency, community support, and long-term maintainability of the AI software ecosystem.	
5.7-8	*Lightweight Performance Profiling* The procured infrastructure shall provide lightweight performance profiling capabilities that can be activated by users on a per-job basis. Profiling data shall be collected at process, job, and node levels using scalable aggregation methods. Data retention times may vary depending on the granularity level. The profiling technology shall impose minimal overhead on application performance. A baseline set of metrics shall be collected transparently, without requiring users to modify or link their applications against specific profiling libraries. The system shall provide monitoring for the hardware components such as: • CPU utilization metrics, including load averages, instructions per cycle (IPC), and instruction mix information. • Memory metrics such as memory footprint, cache utilization (hits and misses), Translation Lookaside Buffer (TLB) hits/misses, load/store operations, and memory interface utilization. • For systems featuring multiple memory types and hierarchical memory tiers, metrics shall be collected separately for each memory type and tier. • I/O subsystem metrics, including number of reads and writes, and read/write bandwidth. • Network metrics, including number of packets, reads, writes, and packet or segment sizes relevant for data movement across high-speed interconnects. • Accelerator utilization metrics, capturing usage statistics for GPUs integrated in the system. The profiling solution shall support flexible extensibility, enabling the addition of new observables and metrics as needed, with the possibility of increased overhead depending on the profiling detail level.	MRQ

	Possible profiling solutions include, but are not limited to, NVIDIA Nsight Systems, NVIDIA Nsight Compute, TensorBoard profiling tools, and other open-source performance monitoring frameworks.	
5.7-9	*Optimized AI Numerical and Tensor Libraries* • The Candidate shall provide highly optimized libraries offering API-compatible implementations for core numerical and tensor operations commonly used in AI and machine learning workloads. The solution shall include optimized libraries for linear algebra, tensor computations, and related primitives, suitable for deep learning frameworks. • The Proposal shall include an optimized fast Fourier transform (FFT) library capable of efficient processing on CPUs and GPUs, supporting AI workload demands such as signal processing and convolution operations. Examples of such libraries include NVIDIA's cuBLAS, cuFFT, and TensorRT. Equivalent open-source alternatives are also acceptable.	MRQ
5.7-10	*Parallel debugger* The software package shall include an advanced debugger and profiler specifically designed to support AI and machine learning workloads, enabling efficient debugging and performance analysis of parallel and distributed AI applications. The provided tools must support heterogeneous architectures commonly used in AI systems. Examples of such tools include NVIDIA Nsight Systems and Nsight Compute but equivalent open-source alternatives are also acceptable.	MRQ

6 Cost Performance Analysis

6.1 Introduction

This section describes the rationale behind the Cost Performance Analysis (CPA) methodology applied for this procurement to provide a fair framework for evaluating the solutions proposed by the Candidate.

The CPA provides a framework that aims to define a value-for-money metric. Higher is the metric, higher is the value of the offered solution, according to defined metrics, for the given capital investment.

This Chapter is organized as follows:

- In Section 6.2 the cost and capacity analysis is reported.
- In Section 6.3 the application performance analysis is described.
- In Section 6.4 the Cost Performance Analysis and related rules is described.

6.2 Cost and capacity analysis

6.2.1 Purpose and scope

The objective of this analysis is to provide a transparent and technology-agnostic framework that enables a fair comparison among solutions by combining three dimensions:

- the throughput of the offered Scale-out AI/HPC subpartition, expressed as the number of accelerators offered;
- the total theoretical peak performance of the offered Scale-out AI/HPC subpartition, expressed as dense FP8 performance;
- the energy-related operational cost expected over the reference lifetime of the offered Scale-out AI/HPC subpartition.

This analysis is intended as an evaluation tool and does not represent a full Total Cost of Ownership (TCO) model; it focuses on the components that are both (i) dominant in operational expenditure for this class of systems and (ii) directly tied to the technical characteristics of the proposed architectures.

6.2.2 TCO approximation and energy OPEX model

The total cost of ownership is calculated according to the following equation:

$$TCO = CAPEX + OPEX$$

For the purposes of the evaluation, the Procurer considers the following simplified approximation of the TCO:

$$TCO = CAPEX_{(aq)} + OPEX_{(en)}$$

Where:

- $CAPEX_{(aq)}$ is the capital expenditure for system acquisition,
- $OPEX_{(en)}$ is the energy contribution to operational expenditure.

For the CPA evaluation, $OPEX$ is approximated to its energy component $OPEX_{(en)}$ in order to: (i) keep the model understandable and reproducible, (ii) reduce the number of subjective variables, and (iii) reward solutions with lower energy footprint.

The Procurer estimates $OPEX_{(en)}$ assuming an average load factor (to reflect typical production operation rather than peak benchmarking conditions), and accounting for the fraction of power dissipated through Direct Liquid Cooling (DLC) versus air cooling. The energy OPEX is computed as:

$$OPEX_{(en)} = PC_p \times LF \times LT \times EP \times (x \times PUE_{DLC} + (1 - x) \times PUE_{AC})$$

Where:

- PC_p= Peak power consumption of the Scale-out AI/HPC subpartition under maximum power operating conditions (High Performance Linpack).
- x = Fraction of the total system power dissipated using DLC technology.

To calculate $OPEX_{(en)}$, CINECA will provide the following values:

LF	Load Factor applied to scale the power consumption to normal operating conditions	0.65

LT	Lifetime of the system in hours (4 years + 1 leap year)	43824
EP	Energy Price in Euro/kWh	0.20
PUE_{DLC}	Estimated yearly Average Power Usage Effectiveness of DLC-based system in CINECA data centre.	1.1
PUE_{AIR}	Estimated yearly Average Power Usage Effectiveness of Air-based system in CINECA data centre.	1.4

6.2.3 Definitions of performance and efficiency quantities

To ensure a consistent comparison across all offers participating in the procurement procedure, the CPA_{cost} uses the following quantities for the Scale-out AI/HPC sub-partition:

1) **Throughput**: where N_{GPU} is the number of accelerators (GPUs) offered in the Scale-out AI/HPC sub-partition which is the most relevant partition of the system.

$$T = N_{GPU}$$

2) **Total theoretical peak performance (dense FP8)**: where P_{FP8} is the theoretical peak performance of a single GPU expressed in TFLOPs for dense FP8 operation.

$$P = P_{FP8} \times N_{GPU}\ [TFLOPS]$$

3) **Energy cost over lifetime**: computed using the energy OPEX model defined above.

$$C = OPEX_{(en)}\ [€]$$

6.2.4 Normalization across offers

Since the above quantities have different units and scales, the CPA_{cost} applies a normalization across all admissible offers. For each offer j, the normalization is computed by comparing its value with the best value observed among all offers k.

- For metrics to be maximized, namely Throughput and Theoretical Peak Performance, the normalized scores are computed as follows:

$$\widehat{T_j} = \frac{T_j}{max_k(T_k)} \qquad \widehat{P_j} = \frac{P_j}{max_k(P_k)}$$

where:

- T_j is the Throughput of offer j, expressed as the number of GPUs offered in the Scale-out AI/HPC sub-partition;

- P_j is the total Theoretical Peak Performance of offer j, expressed as aggregate dense FP8 theoretical peak performance;
- k spans all admissible offers.

For the metric to be minimized, namely Energy Cost, the normalized score is computed as follows:

$$\widehat{C_j} = \frac{min_k(C_k)}{C_j}$$

where:

- C_j is the Energy Cost of offer j over the reference lifetime;
- k spans all admissible offers.

This normalization produces scores in the interval $(0, 1]$, where 1 denotes the best value among the compared offers for the considered metric. In particular, for the Energy Cost score, a lower energy cost results in a higher normalized score.

6.2.5 Candidate obligations (data to be declared)

For the CPA_{cost} computation, each candidate shall declare at minimum:

- N_{GPU} for the Scale-out AI/HPC sub-partition. If no GPUs are offered, the CPA value will be considered 0.
- P_{FP8} (TFLOPs, dense FP8) for the proposed GPU model.
- PC_{peak} of the Scale Out AI/HPC sub-partition configuration (kW) at maximum power operating conditions (worst case scenario as for High Performance Linpack).
- cooling split x (fraction of power dissipated via DLC) for the platform solution.

6.2.6 Cost and capacity analysis score

The CPA_{cost} score is built by combining two complementary components:

1. a *Performance Score*, which measures the overall capacity of the offered Scale-out AI/HPC sub-partition in terms of system size and aggregate theoretical peak performance;
2. a *Balance Score*, which measures whether the offered system provides a coherent relation between system size and theoretical peak performance, while taking into account the energy cost required to achieve such configuration.

The CPA is intended to reward offers that provide high computational capacity, while also favouring solutions that achieve such capacity with a balanced system profile and a lower energy-related operational cost.

The *Performance Score* is computed as the simple arithmetic mean of the normalized Throughput score and the normalized Theoretical Peak Performance score:

$$PS_j = \frac{\widehat{T}_J + \widehat{P}_J}{2}$$

The *Balance Score* is computed as follows:

$$BS_j = \left(1 - \left|\widehat{T}_J - \widehat{P}_J\right|\right) \times \widehat{C}_J$$

where:

- $\widehat{T}_J$ is the normalized Throughput score;
- $\widehat{P}_J$ is the normalized Theoretical Peak Performance score;
- $\widehat{C}_J$ is the normalized Energy Cost score.

The raw CPA score for offer j is computed as a weighted mean of the *Performance Score* and the *Balance Score*:

$$CPA_j^{raw} = \frac{\omega_{PS} \times PS_j + \omega_{BS} \times BS_j}{\omega_{PS} + \omega_{BS}}$$

Where:

- PS_j is the *Performance Score* of offer j;
- BS_j is the *Balance Score* of offer j;
- ω_{PS} is the weight assigned to the *Performance Score*;
- ω_{BS} is the weight assigned to the *Balance Score*;

For this procurement, the weights are defined as follows:

$$\omega_{PS} = 0.85 \quad \omega_{BS} = 0.15$$

Since the weights sum to one, the formula can be written equivalently as:

$$CPA_j^{raw} = \left(\omega_{PS} \times PS_j + \omega_{BS} \times BS_j\right)$$

The raw CPA score is bounded in the interval $[0,1]$, where higher values indicate a better value.

To ensure comparability across all admissible offers, the final CPA_{cost} score used for evaluation is obtained by normalizing the raw CPA_{cost} scores with respect to the best-performing offer and scaling them to a 25-point scale.

$$CPA_{cost} = 25 \times \frac{CPA_j^{raw}}{max_k(CPA_k^{raw})}$$

where:

- k spans all admissible offers.

6.3 Application performance analysis

6.3.1 Benchmark suite

The benchmark suite is composed by the following training tasks, which are a subset of the publicly available MLPerf training benchmark suite[1]

#	Benchmark	Partition	Short description	website
1	**Training Benchmark:** **-Retinanet** **-Stable diffusion v2** **-GPT-3**	Scale-out AI/HPC Sub-partition.	Benchmarks that represent a challenge for the targeted system configuration and that cover a diversity of user workloads. The 3 training tasks are a subset of the available MLPerf training benchmark suite.	https://github.com/mlcommons/

Table 3: Benchmark application composing the benchmark suite.

6.3.2 Benchmark execution

Datasets, instructions, and rules on how to configure, set-up, and run the benchmarks are all available in the websites referred in Table 3.

6.3.3 Benchmark analysis report

To evaluate the benchmark results, a benchmark analysis report provided by the Candidate is required. The report will assess the performance, in terms of the selected benchmark applications, of the offered solution. Given that AI is one of the main drivers, results for MLPerf training benchmarks are expected to be reported.

6.3.4 Application performance analysis score

To assess the quality of the offer, a benchmark analysis provided by the Candidate is required. To support this evaluation, each Candidate will provide the time-to-train latencies (in seconds) projected on the offered solution and the information provided in the Technical Response Template document.

For each Candidate a value $CPA_{app.perf}$ is assigned to the benchmark analysis report.

The application performance score for each Candidate is therefore calculated with:

$$CPA_{app} = 10 \times \frac{CPA_{app.perf}}{max_k(CPA_{app.perf})}$$

where:

[1] https://mlcommons.org/benchmarks/training/

- k spans all admissible offers.

6.4 Cost Performance Analysis evaluation

For each Candidate C, the cost-performance analysis coefficient CPA_C will be evaluated according to the following formulas:

$$CPA_C = CPA_{cost} + CPA_{app}$$

$$CPA_C^{norm} = \frac{CPA_c}{max_C(CPA_c)}$$

The score will be calculated with:

$$Score = MAX.points * CPA_C^{norm}$$

where $MAX.points$ is the maximum points associated to the CPA score: 35 points.

The Candidate can design the offered solution to provide the optimal *CPA* score. This configuration must be replicable during the configuration of the system with the tools and technologies provided in the offer and available to CINECA system administrators.

7 Maintenance and infrastructure availability

The Offer for the procured infrastructure must include a maintenance and support service that ensures high availability and stability of the procured infrastructure. In the following with the term Supplier, it is referred the Candidate awarded for this procurement.

7.1 Maintenance and support requirements

Req.	Description	Category
7.1-1	*Maintenance and support duration* The Candidate will offer maintenance and support of the offered infrastructure according to the procurement document "Draft contract acquisition".	MRQ
7.1-2	*Maintenance and support coverage* The maintenance and support will cover all key hardware and software (incl. firmware and all offered programming environment software) components of the procured infrastructure. This includes all infrastructure component (e.g., racks and power supplies) and network components except for those components provided by CINECA. The Candidate will describe all components not covered by the system maintenance and support. The customer maintenance and support times may be restricted to normal working hours. At least the standard working hours on all working days (excluding weekends and Italian public holidays), i.e., 8×5, will be covered. The Supplier must ensure the provision of the maintenance and support services, even if there is a dispute with CINECA.	MRQ
7.1-3	*Special software support coverage* Software, whose malfunction could harm the system stability and hardware health, will be supported by the Supplier. For example, if power capping techniques integrated in the workload manager are used for system operation, these workload manager capabilities must be fully supported by the Supplier.	TRQ
7.1-4	*Reaction times* The Supplier guarantees an appropriate reaction time upon hardware and software issues.	MRQ
7.1-5	*On-site stock* The Supplier will populate and maintain an on-site stock in CINECA's facility to ensure the availability of replacement parts, especially for components whose loss significantly affects system availability or utilization. However, the Supplier may rely on a facility other than CINECA's for stocking spare parts near CINECA's data centre (3 hours). The Candidate will provide a list of the intended spare parts included in the on-site stock.	MRQ

7.1-6	*Support* The Supplier will include one full time equivalent (FTE) position, based on one or multiple qualified persons, to ensure system support during working hours. The personnel will support CINECA primarily in failure analysis, hardware support (including spare and replacement part logistics if necessary) and software support for cluster management and system management software. If the FTE is based on multiple persons, a reasonable team size must not be exceeded, and an appropriate coverage of the relevant support fields must always be ensured.	TRQ
7.1-7	*Preventive maintenance and early errors' detection actions:* The Supplier will perform preventive maintenance and early detection of errors actions to replace components that are likely to fail soon. Example of possible preventive maintenance actions are the replacement of components (disks, networking equipment) based on error counter information prior to the point where system operation is impacted and the replacement of memory components exhibiting high single-bit error rates prior to the occurrence of a (fatal) double-bit error. The Candidate will document these actions included in the Offer.	MRQ
7.1-8	*Data deletion* The Supplier will ensure that all client data stored on any, user accessible, non-volatile storage component (incl. SSD, HDD) are deleted when components are taken off-site as part of the system maintenance. Data deletion may occur off-site but must conform to common data protection guidelines. Alternative means that ensure the confidentiality of the data stored on non-volatile storage components (e.g., destruction by the customer) may be proposed.	MRQ
7.1-9	*Serviceability constraints* The Candidate will document all serviceability constraints affecting system availability. Examples of such serviceability constraints include sibling nodes that must be taken offline for a node replacement or rack components that need to be taken out of service for network servicing.	MRQ
7.1-10	*Escalation management process* The Supplier will provide an escalation management process to manage problems priority and critical/non-critical issues.	MRQ
7.1-11	*Regular maintenance and support meetings* It may be required to host regular face-to-face meetings (up to eight per year) to discuss the status of the installation and address any issues. The Supplier shall ensure the availability of the necessary travel funds to allow the attendance of key support personnel at these meetings.	MRQ
7.1-12	*Pre-production qualification acceptance*	MRQ

	The Supplier will declare whether the system design and maintenance concept enable the system to pass the acceptance tests proposed in Chapter 8.	
7.1-13	*Responsibilities and roles* The Candidate will describe the roles and responsibilities of all parties involved during system operation in the form of a RACI- (Responsible, Accountable, Consulted, Informed)-model.	MRQ
7.1-14	*Security patches and software updates* The Supplier will make security patches for all supported software components available for installation in an adequate period following the release of the component by the vendor. Availability of Security Updates: All offered system components must receive security updates throughout the lifetime of the system. In case of (disclosed) major vulnerabilities, especially those allowing privilege escalation, the supplier will ensure the immediate collaboration to place mitigations and/or patches. Furthermore, the Supplier will ensure the release and application of software updates (firmware, drivers, micro-codes) for bug fixing or adding new features; note that the term "update" also refers to new versions ("releases") of the software. Besides software updates the supplier will provide tested "recipes" for update processes, to ensure the whole system stability and reduce the possible services downtime or degradation.	MRQ
7.1-15	*Maintenance and support service regulation* The maintenance and support service shall encompass all necessary activities to ensure compliance with applicable regulatory requirements for software and equipment, including adherence to all relevant European, national, and regional laws and regulations. All goods covered by the service at its commencement including repaired or replacement parts must conform to current regulations and any subsequent updates. All maintenance and support interventions shall be thoroughly documented. The Supplier or its authorized agents are required to deliver the necessary technical assistance strictly in accordance with the conditions and response times specified in the contract. The Supplier is responsible for the professionalism and qualifications of the technicians assigned to the service. All replacement parts supplied must carry the CE marking and comply with prevailing technical and safety standards, including any future regulations, particularly those issued by UNI (Italian National Standards Body) and CEI (Italian Electrotechnical Committee) if applicable.	MRQ
7.1-16	*Maintenance periodic reporting* Periodic maintenance and maintenance activities must be reported, including: • Ticket lists issued by the call center, including the relevant details.	MRQ

	• List of technical assistance interventions detailing the activities carried out and the total duration of the disruption. • Reports of possible preventive maintenance interventions. • Analysis of repeated failures. • Conformance ratios to SLAs. Candidate will describe this periodic reporting in the Offer.	
7.1-17	*Documentation requirements* The Candidate will describe the support workflow for software and hardware failures, including information about replacement part logistics and SLAs.	MRQ

7.2 Professional services

Req.	Description	Category
7.2-1	*Professional services for networks* The Supplier shall provide comprehensive professional services for the network infrastructure described in Chapter 5.2, covering the setup and tuning phases as well as the production and operational phases. The support shall be guaranteed for the entire lifetime of the System (60 months). All the required services shall be clearly described and included in the Offer. The professional services shall include, at minimum: • Design, installation, and configuration services for all network devices and equipment included in the Offer, ensuring full compliance with the technical specifications and the required operational architecture. • Post-implementation professional support services amounting to at least 15 (fifteen) days, which may also be delivered remotely. These services shall support activities such as the implementation of additional network configurations not foreseen in the initial design, functional analysis, troubleshooting of operational anomalies, and performance optimisation. • Training services for CINECA personnel on the technological architecture, configuration, and operational management of the supplied network equipment. The training shall be delivered through dedicated sessions limited to a maximum of 10 participants per session. The Supplier shall describe the training plan in detail, including number of training days and proposed session structure. Furthermore, the Candidate must be a certified partner or authorised reseller of the manufacturers of the equipment included in the Offer, thereby demonstrating the capability to successfully perform all activities required under these Technical Specifications.	TRQ

7.2-2	*Professional services for OpenStack platform* The Supplier shall provide comprehensive specialist support for the design, deployment, tuning, and lifecycle maintenance of the OpenStack cloud platform. The Offer shall include: • *a minimum of 30 (thirty) days* of specialist support dedicated to the definition of the OpenStack architecture, the deployment of its services, and the initial performance tuning activities; • *a minimum of 4 (four) days per year* of specialist support for the entire operational lifetime of the OpenStack environment, covering corrective and evolutionary maintenance, troubleshooting, configuration optimisation, and technical advisory. Such support shall encompass advanced consultancy services, issue resolution, service tuning, updates and upgrades, and technical assistance to the Beneficiary's personnel throughout the lifecycle of the OpenStack deployment.	TRQ
7.2-3	*Professional services for Orchestrated computing partition platform* The Supplier shall provide comprehensive specialist support for the design, deployment, tuning, and lifecycle maintenance of the Orchestrated computing partition platform. This environment shall be deployed and managed using industry-standard automation and orchestration tooling, including but not limited to Kolla-Ansible–based workflows where applicable. The Offer shall include: • *a minimum of 20 (twenty) days* of specialist support dedicated to the definition of the Kubernetes architecture, the deployment of its components and services, and the initial performance tuning activities; • *a minimum of 4 (four) days per year* of specialist support for the entire operational lifetime of the Kubernetes environment, covering corrective and evolutionary maintenance, troubleshooting, configuration optimisation, and technical advisory. Such support shall encompass advanced consultancy services, issue resolution, cluster tuning and optimisation, updates and upgrades, validation of high-availability and resilience configurations, and technical assistance to the Beneficiary's personnel throughout the lifecycle of the Kubernetes deployment.	TRQ

7.3 Licenses

Req.	Description	Category
7.4-1	*Licenses*	MRQ

	Where applicable, the Candidate must provide licenses for all offered software for the complete duration of the maintenance and support time frame.	
7.4-2	*Licenses' list* Candidate must provide the complete list of all applicable licenses provided with their quantities.	MRQ

7.4 Infrastructure availability

The procured infrastructure must seek the highest availability to the end users. CINECA will report on monthly availability for the server nodes (*Availability_c)* and for the storage (*Availability_s)*. They are calculated independently with the following formulas:

$$Availability\_c = \frac{\sum_i^N a_i}{\Delta T \cdot N - \sum_m \sum_i^{N_m} d_i}$$

Where:

- a_i is the availability of each single node "i" (front-end and Compute Nodes). Availability here means that the node is up and running, reachable by the users directly or through the WLM slurm or the cloud dashboard.
- ΔT is the time interval considered (i.e., a month expressed in hours).
- N is the total number of Compute Nodes.
- m denotes a scheduled maintenance intervention.
- N_m is the number of nodes involved in the scheduled maintenance "m".
- d_i is the time that node "i" spent in scheduled maintenance "m". For each node involved in maintenance, the time d_i starts after action of system administrator that manually drains the node.

$$Availability\_s = \frac{\Delta T_a}{\Delta T - \Delta T_m}$$

Where:

- ΔT_a is the time when the storage infrastructure is available to the users.
- ΔT is the time interval considered (i.e., a month expressed in hours).
- ΔT_m is the time when the storage infrastructure is under scheduled maintenance and therefore unavailable to the users.
- m denotes a scheduled maintenance intervention.

For the AI/HPC partitions planned maintenance[2] interventions will be scheduled for up to 7 days per year and every month for a duration of 8 hours.

[2] Therefore, these interventions do not contribute for the calculation of the total Availability_s and Availability_c percentages as expressed in the respective formulas.

Req.	Description	Category
7.5-1	*Targeted monthly availability* The design of the procured infrastructure architecture and the maintenance service must aim for a monthly availability of 99%.	MRQ
7.5-2	*Minimum monthly availability* The design of the procured infrastructure architecture and the maintenance service must aim for a minimum monthly availability of 85%	MRQ

8 Installation and acceptance

8.1 Installation time schedule and project management

The Proposal for the procured infrastructure should include the necessary planning and project management resources for the installation of the system. In the following the term Supplier refers to the Candidate awarded for this procurement.

8.1.1 System Installation

Req.	Description	Category
8.1.1-1	*Project Management* The Supplier will provide project management resources for the system installation.	MRQ
8.1.1-2	*Benchmarking Support* The Supplier will provide expert support for the benchmarking of the system. The optimization of benchmark performance, rule conforming execution and the submission of the results to the official lists will be performed by the Supplier.	MRQ
8.1.1-3	*Installation Time* The timeline of delivery, installation, acceptance and start of operations are detailed in the tender document "Draft contract acquisition".	MRQ
8.1.1-4	*Best practices and security* • During installation the system will be accessible only through CINECA VPNs or Bastion hosts. • All the administrative user's passwords of all the installed equipment must be changed from their default values in the very early stages of deployment. • The used passwords must be adequately strong. • The passwords of admin users will be disclosed only to the essential staff. • Every person that doesn't need to know a certain password to operate will be allowed to admin accounts by passwordless and/or ACL mechanisms. • Every person that doesn't need admin access to a given equipment will be allowed with unprivileged user. • The secrets used for the management cluster and services must be different from the ones used for regular production nodes. • The secrets must not contain common substrings. The secret databases must be protected by adequate passwords. • Directory services must not expose passwords (querable) and all the secrets must be stored in "hashed" form.	MRQ

	• The Service node's cluster networks must be segregated and only the essential services must be "published" to the regular nodes cluster networks. • The out-of-band management networks must be segregated and connected only to the management Service nodes. • Early access users, benchmarkers and all the people not involved in the installation process can access the system only via login nodes. • No shell enabling access to the Service or management nodes will be allowed through connections coming from regular cluster nodes. • A clear operational recipe must be made available to change the passwords and the secrets of all the services and nodes, as well as adequate automated helper procedures. • All the inactive and unused accounts and related secrets must be closed and/or deleted immediately. • All data in storage resources must be protected and disclosed only to the essential people. • In case of teams/subcontractors/main contractor handover during installation all the secrets and accounts must be contextually changed. • During acceptance handover all the secrets must be changed. • All the system logs during installation must be collected and aggregated using effective techniques to allow queries and forensic analysis.	

8.1.2 Supply and installation project

Req.	Description	Category
8.1.2-1	Installation project plan The Supplier is responsible for creating a supply and installation project for the procured components. This project must detail the delivery times of the various parts of the system, including any downtime that might affect the operation of CINECA's infrastructure. The project must include: • Details of the offered configuration and integration in the CINECA's system architecture, including setup and interconnection schemes. • Details of hardware and software installation plan, configuration and optimization of the components and partitions. • Details on how the interaction with CINECA staff is organized and foreseen. • Implementation plan for procured infrastructure acceptance (see Section 8.2). All interactions with CINECA staff, all training activities, as well as the documentation produced within the project, can be in Italian or English.	MRQ

8.1.2-2	*Time Schedule* The Candidate will describe the time schedule for the system installation in detail and in terms of a GANTT chart. The time schedule will provide expected dates for the production and delivery of system components, installation, bring-up and acceptance of the procured infrastructure.	MRQ
8.1.2-3	*Project Risks* The Candidate will provide a list of risks that could negatively affect the installation and early operation of the procured system. For each risk, the Candidate will give an indication of likeliness, provide a description of the expected impact and risk mitigation measures that will be implemented as part of the contract.	MRQ
8.1.2-4	*Responsibilities and Roles* The Candidate will describe the roles and responsibilities of all parties involved during system installation and early operation in the form of a RACI- (Responsible, Accountable, Consulted, Informed)-model.	MRQ

8.2 Acceptance procedure

8.2.1 Documentation requirement

The Candidate will declare whether he agrees to the acceptance tests defined in this Section (with details specified in accordance with the Offer). Please note that all acceptance tests verifying committed functionality and performance values included in the Offer are non-negotiable.

8.2.2 Execution of acceptance tests

All acceptance tests will be performed by the Supplier together with, or directly informing, CINECA staff.

For the acceptance procedure the following rules apply:

Req.	**Description**	**Category**
8.2.2-1	*Acceptance rules* The Compute Nodes will run the full operating system stack. The latest security updates will be installed. The system may not be reconfigured for different benchmarks unless this process is fully integrated in the workload manager and will be available at user-level during the production phase of the procured infrastructure. The benchmark runs will be performed using the offered compiler suite. If the Offer includes multiple compiler suites, the Supplier may choose a	MRQ

	different combination for each benchmark. All tests will be performed using the production workload manager.	

8.2.3 Provisional acceptance tests

8.2.3.1 Hardware checklist

Req.	Description	Category
8.2.3.1-1	*Hardware checklist* Completeness and consistency of the delivered and installed hardware will be checked against the Offer.	MRQ
8.2.3.1-2	*Failure thresholds* The thresholds for defective components described in the following must not be exceeded. For equipment not listed below no fatal deficiencies may exist for the provisional acceptance test to be passed. • Compute nodes: less than 2% of nodes may be not functional. • Service nodes: all nodes must be functional. • Network links: less than 0.1% of the links may be not functional.	MRQ

8.2.3.2 Software Checklist

Req.	Description	Category
8.2.3.2-1	*Software checklist* Completeness and consistency of the delivered and installed software will be checked against the Offer. All components must be installed for the test to be passed.	MRQ

8.2.3.3 Functional Tests

Req.	Description	Category
8.2.3.3-1	*Acceptance plan* The Supplier in agreement with CINECA will provide an acceptance plan to verify, with a series of functional tests, the suitability of the components against the expected performance level reported in the Offer. All the components (hardware and software) must be checked against their performance level as described in the	MRQ

	Offer. Components may be grouped together and verified with a single test in accordance with CINECA staff.	

The list of the functional tests included in the acceptance plan will ultimately depend on the system design and the Offer. For the benefit of the reader a non-exhaustive list of tests may include:

- Verification of the power and cooling infrastructure.
- Verification of power management system.
- Verification of health checks and monitoring in accordance with the Offer.
- Verification of cluster management including management network (collection of metrics, redundancy, node reinstallation and configuration).
- Verification of data network (reachability of Compute Nodes and service nodes, bandwidth, and latency performance).
- Verification of the stability of the system software, firmware and hardware (component stress tests, see Section 8.2.3.4).
- Verification of single functional component performance level (node, group of nodes, rack, group of racks).
- Verification of system level performance (Benchmark suite, IO partition, see Sections 8.2.3.5-8.2.3.7),
- Verification of software specifications and offered features.
- Verification of other commitments made in the Offer.

8.2.3.4 Stress tests

The following synthetic tests are typically performed to stress the hardware components. In case of failure, the faulty components must be replaced, and the test will be restarted on the affected component. All these tests must be passed successfully.

- A local, optimized HPL will be run on each single node, in parallel for 30 minutes without failure. The Supplier may suggest an appropriate alternative tool.
- A memory stress test will be performed on the system. CINECA proposes a modified STREAM version which uses >95% of the system memory for this test. The Supplier may suggest an appropriate alternative tool.

8.2.3.5 Application Benchmarks

The benchmarks included in the benchmark suite (see Section 6.3) will be executed with the baseline values as provided by the Supplier in the Offer. For the test to be passed, all committed benchmark results must be achieved and assessed by the Procurer in accordance with the Candidate.

8.2.3.6 Rules for HPL Benchmarks

The HPL benchmark will be executed on the CPU and GPU partition separately, according to the TOP500 list rules. During the LINPACK benchmark, the power consumption will be measured according to the GREEN500 run rules.

8.2.3.7 I/O Performance

The I/O performance of the filesystems shall be measured using IOR; however, the Supplier may propose an alternative tool, subject to mutual agreement between the parties. The committed I/O performance must be achieved within the relative tolerance of 2%.

8.2.4 Pre-production qualification

The stability of the system will be tested over the course of one month under near-production conditions. For this purpose, the system will be filled with an arbitrary, well behaving, workload (i.e., a workload that does not trigger out-of-memory situations or other software exceptions). In this phase an early access can be provided to selected and experienced users.

Req.	Description	Category
8.2.4-1	*Pre-production availability* The Supplier will replace failed components and tune the infrastructure configuration during the pre-production phase to reach at least one week with an availability that must result 90% or above.	MRQ

8.2.5 Final acceptance

The final acceptance will validate the proper functioning of the entire system after the preproduction qualification period.

9 EU added value

Req.	Description	Category
9-1	*Reinforce the EU digital technology supply chain* The Tenderer shall provide a detailed description of how the proposed system contributes to strengthening the digital technology supply chain within the European Union, encompassing both hardware and software aspects. Examples may include, but are not limited to, utilizing EU-based manufacturing facilities for system components, supporting the development of EU-originated Intellectual Property, and fostering collaboration with EU technology partners.	TRQ
9-2	*Level of integration EU technologies* The Tenderer shall provide a comprehensive description of the extent to which European technologies are integrated into the proposed system. This includes, but not limited to, the adoption of existing or forthcoming R&D outcomes derived from EU-funded research programs or initiatives supported by the R&D programs of EuroHPC participating States.	TRQ
9-3	*Collaboration* The Tenderer shall provide a comprehensive description of the level of alignment the proposed collaboration topics have with the objectives and priorities of EuroHPC and the Hosting Entity.	TRQ

10 Documentation

All documents relevant to the bid shall be submitted in machine-readable electronic format.

Req.	Description	Category
10-1	*Installation plan documentation* The Tenderer shall provide comprehensive documentation encompassing all phases of installation and maintenance for the respective systems.	DOC
10-2	*Milestones' documentation* Upon completion of each installation project milestone, the following documentation shall be submitted in PDF format: • Technical datasheets and manuals for all installed components. • Updates to any existing manuals reflecting adaptations related to newly installed or upgraded components.	DOC

10-3	*Documentation after installation*
	
	The Tenderer commits to delivering comprehensive digital documentation upon completion of the installation, provided in PDF format. This documentation shall include:
	
	• A general description of the solution components.
	• Diagrams illustrating the finalized system connections.
	• Detailed hardware procedure instructions covering:
	  ○ Start-up,
	  ○ Shutdown,
	  ○ Diagnosis and troubleshooting in case of hardware issues.
		DOC

Annex 1: system glossary

Term	Description
Backbone	Site-wide Ethernet network (40GbE or 100GbE)
CINECA	Interuniversity Consortium
DDR	Double Data Rate
DIMM	Dual In-line Memory Module
HPL	High Performance Linpack (see top500.org)
CPU	Central Processing Unit
GPU or GPGPU	Graphic Processing Unit or General-Purpose Graphics Processing Units usable for computation
HA	High-Availability. Mechanism to ensure service availability in case one of a component failure
HPC	High-Performance Computing
LACP	Link Aggregation Control Protocol
NVM	Non-volatile memory
PCIe	Peripheral Component Interconnect Express
PDU	Power Distribution Unit
POSIX	Portable Operating System Interface for Unix
RAID	Redundant Array of Inexpensive Disks. Mechanism to prevent from disk failures by storing redundant information on additional disks (mirror, parity...)
SDRAM	Synchronous Dynamic Random Access Memory
SR-IOV	Single-root input/output virtualization
UPS	Uninterruptible Power Supply
VM	Virtual machine

RHEL	Red Hat Enterprise Linux
CDU	Cooling Distribution Unit
MN	Master nodes
SN	Service Nodes
CN	Compute Nodes, either being CPU- or GPU-based nodes
VN	Visualization Nodes
LN	Login Nodes
TOR	Top Of Rack
HPF	High Performance Ethernet Network
LTS	Long-Term data Storage
LLM	Large-Language Model
NIC	Network Interface Card
HCA	(InfiniBand) Host Channel Adapter
Core	Set of integer and floating calculation units managed by a control unit and capable of executing operations between internal registers and/or external memory. A single Processor may consist of several Cores.
Socket	Connector used to interface a Processor with a motherboard.
Processor	Execution unit constituted by one or more Cores and able to execute a portion of computation independently from the other Processors. Typically, a Processor is constituted by a single chip connected to the central memory and other hardware devices of the system via a single Socket.
Device	Execution unit that performs specific computational or communication tasks to aid the processor in carrying out the execution of a process. Examples include graphics processor units, cards that offer acceleration for floating point intense workloads, other forms of co-processors, network interface cards and storage cards.
Node	Set of Processors, memory areas and Devices. The Processors of a single Node access a shared memory address space through load/store instructions. Devices may feature a separate address space.

Compute Node	A Node dedicated to compute workloads. Nodes of CPU and GPU partition are considered Compute Nodes. Those Nodes are typically managed by the Workload Manager.
CPU Node	A Compute Node that is part of the CPU partition based only on CPUs for data processing.
GPU Node	A Compute Node that is part of the GPU partition based with CPUs and GPUs for data processing.
Front-end Node	A Node dedicated for user access, software and data management. Login Nodes and Visualization Node are considered Front-end Nodes.
Login Node	A Node used by users to submit jobs in the System and to compile applications.
Visualization Node	A Node designed and used specifically for visualization workloads.
Management Node	A Node used for system management. Nodes of Service and Master partition are considered Management Nodes.
Service Node	A Node used for running specific system services (e.g., workload scheduler). A supercomputer may need many Service Nodes.
Master Node	A node used to System Administrator to manage the System. On these nodes are contained tools and applications used to manage the System.
Interconnect	Devices and apparatus that implement a network of Nodes featuring low communication latency and high bandwidth. Typically, all Compute Nodes, Login Nodes and potentially other Nodes are integrated in the Interconnect. The Interconnect hardware is accompanied by appropriate software components to enable message passing between processes on different Nodes. In addition, the Interconnect may integrate storage systems
Filesystem	Technology to manage non-volatile storage components by means of a file abstraction. The file-system technology may be compliant with (official) standards such as POSIX. Examples include XFS, Ext4, IBM Spectrum Scale, Lustre, NFS and pNFS.
Parallel Filesystem	Filesystem accessible in a shared context through a network (potentially the Interconnect) that ensures global consistency (with specific implementation-dependent semantics) of the address space.
Swap	Space on disk (or comparable non-volatile storage components) used by the Operating System for memory paging.
Tiered Storage Solution	Storage solution based on different storage technologies, which are presented as a unique file namespace. The system provides an automatic procedure of data migration across different tiers (types) of storage devices and media.

Batch System	Software component responsible for the management and the scheduling of resources (Nodes) and interactive or batch jobs.
Resource Management System	Software component responsible for the launch, execution and teardown of batch jobs on Nodes.
Workload Manager	Software component consisting of the combination of a Batch System and the Resource Management System

Table 4: List of acronyms and common terms.

--End of Document—