EuroHPC
Joint Undertaking

European High Performance Computing

Joint Undertaking

Call for tenders EuroHPC/2024/OP/0007 -

ACQUISITION, DELIVERY, INSTALLATION AND HARDWARE AND SOFTWARE MAINTENANCE OF THE UPGRADE OF THE EUROHPC SUPERCOMPUTER LEONARDO – LISA

Open procedure

TENDER SPECIFICATIONS

Part 2: Technical Specifications

September 2024

Ver 2.1

Table of Content

1. Preliminary Information concerning the invitation to Tender

1.1 Tender Overview

CINECA has a proven history in providing supercomputers at the top of the most powerful systems in the world as ranked in the top500.org list. The strong partnership with EuroHPC initiative led to the procurement and provisioning of Leonardo, a EuroHPC Joint Undertaking owned system, that constitutes a significant step forward in elevating the European scientific research. The system was ranked fourth in the November 2022 top500.org list and is currently ranked as the seventh most powerful supercomputing system at the time of writing. Leonardo reinforced the European sovereignty in High Performance Computing that is considered strategic asset for enabling the technological growth of the Union member states. The large increase in computing demand – thanks among others to the launch of new flagship European initiatives tackling the most advanced computational challenges – ranges over all levels of computing, forcing to continuously provide researchers with the most advanced computing technologies.

This project aims for an upgrade of Leonardo supercomputer. The authors acknowledge the emerging pivotal role of artificial intelligence (AI), and the new capabilities reached for example by Large Language Models, multi-modal generative AI, with applications in all fields of human knowledge. The project is codenamed LISA - Leonardo Improved Supercomputing Architecture - from the Leonardo masterpiece Mona Lisa.

1.2 Description of the procurement

The targeted system architecture is designed to address new evolving user needs involving AI workloads in the user workflows. In conjunction with the HPC capacity of Leonardo, LISA will offer an AI-optimised partition, complementing the computing service portfolio of the whole infrastructure. Moreover, while this project targets a considerable extension of the capability of Leonardo supercomputing system, a strong emphasis is devoted to reducing the impact of the upgrade installation on Leonardo normal operations and minimize potential downtimes.

With LISA, EuroHPC Joint Undertaking and CINECA aims therefore to increase the capability of Leonardo in addressing AI workloads. LISA building blocks confirm the technological foundations laid out in the Expression of Interest (EoI). They include next-gen components such as state-of-the-art GPUs, high memory bandwidth and CPUs. By adopting cutting-edge processing units, LISA significantly enhance low-precision performance, that is critical for AI workloads since they often operate at FP16 or lower. LISA will multiply the AI computing capacity available to users. Moreover, LISA's infrastructure will be designed to host AI models and user data efficiently. LISA architecture will leverage fast access, high bandwidth memory, shared among multiple tightly interconnected (Graphical) Processing Units within a NUMA-like node. This configuration enables seamless handling of complex AI models with billions (and even trillions) of parameters, such as wide spreading Large Language Models. Notably, this approach minimizes performance impact due to tiered data storage.

To support large-scale AI model training, LISA's network must be highly performant. The infrastructure design must enable effective data exchange across NUMA nodes. This is crucial for data exchange patterns involved in large AI model training, requiring the highest bandwidth and lowest latency at scale, meaning involving all the processing units concurrently.

Finally, the storage solution must support high IOPS, multi-protocol workloads, and be flexible enough to provide a common Data-Lake layer between LISA and Leonardo, as foreseen in the original EoI proposal. The Data-Lake storage will be provided by CINECA and will not be part of the procured solution for the procedure to focus on AI computing capacity.

We anticipate the upgrade to be adequate to address the recent increase of computing power demand. In particular, the following examples can be cited:

- Request for developing and refining AI models, from various scientific disciplines.
- Request for building high-definition Digital Twin (Destination Earth, Digital twins of the Ocean, Digital Twin of the Human Body, Digital Twin of the complex urban system, etc.)
- Request for EU-Flagship project and scientific challenges (LHC High Luminosity, Square Kilometre Array, etc.)
- Centres of excellence (Cheese2, Excellerat2, MaX, etc.)
- EuroHPC Research and Innovation funded projects
- EuroHPC and ISCRA (Italian Supercomputing Resources Allocation) action for the excellence in science peer reviewed open access
- Weather forecast (ECMWF, Italia Meteo Agency)
- Request from Italian research institutions (new RRF funded National Foundation for HPC, Big data and Quantum Computing)

All these actions need to be supported by providing extensive AI computing resources.

2. Document definitions and glossary

2.1 Tendering procedure definitions

2.1.1 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

2.1.2 Categories of requirements

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

Requirements priority	Requirements category	Description
Mandatory	**MRQ**	*Mandatory Requirements.* These specifications are considered essential for the procured infrastructure and must be fulfilled by all best and final Offers. Mandatory Requirements will be assessed for each offers submitted. The Offers may include improvement with respect to the mandatory requirements, for example in terms of numerosity, capacity, capability. This improvement will be taken into consideration in evaluating the Offers. Final Tenders which will not be compliant with all Mandatory Requirements will be rejected.
Targeted	**TRQ**	*Highly targeted requirements.* These are highly desired specifications for the procured system. In contrast to Mandatory Requirements, failure to provide targeted requirements will not lead to the rejection of the best and final Offer provided by the Candidate.
Mandatory	**DCS**	*Data Centre Specifications.* The Offer must comply with the detailed data center specifications. However, while complying with the requirements' framework, the Candidate is allowed to propose alternatives at its own cost, in order to provide the adequate data center integration of the offered solution into CINECA infrastructure.
Mandatory	**DOC**	Documentation Documentation that must be included in the Proposal. All documentation items are mandatory and must be provided by all Candidates in their Proposal. Documentation requirements will be assessed for each Proposal submitted based on the quality of the response.

Table 2: Categories of requirements

2.2 Glossary

Term	Description
GW	Gateway
Backbone	Site-wide Ethernet network (40GE or 100GE)
CINECA	Interuniversity Consortium
DDR	Double Data Rate
DIMM	Dual In-line Memory Module
HPL	High Performance Linpack (see top500.org)
GPGPU or GPU	General-Purpose computing on Graphics Processing Units. Graphic processing unit 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
NMV	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	Management nodes
SN	Service Nodes
CN	Compute nodes
VN	Visualization nodes
LN	Login Nodes

Table 3: List of acronyms and common terms

Concept	Definition
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. Compute Nodes are typically managed by the Workload Manager.
Login Node	A Node dedicated for user access, software and data management. The extent to which pre- and post-processing workloads are supported on Login Nodes is site specific.
Visualization Node	A Node designed and used specifically for visualization workloads.
Service Node	A Node used for running specific system services. A supercomputer may constitute many Service Nodes.
Management Node	A Node used for system management. A system usually contains one or two Management Nodes.
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 technical definitions

2.3 Unit of measure

Regarding units for memory and storage capacities, the following applies. Unless stated otherwise, SI units (rather than ISO/IEC 80000 prefixes) are used in the technical specifications and should be used for the Proposal. For example:

1 kB = 1000 bytes, 1 MB = 1000 kB, 1 GB = 1000 MB, 1 TB = 1000 GB, 1 PB = 1000 TB

The Proposal should preferably exclusively use SI prefixes. Where this is not possible, the use of IEC (binary) prefixes must be made clearly visible.

The compute performance of a system may be assessed using the following unit:

1 kflop/s = 1000 floating point operations per second (flop/s)
1 Mflop/s = 1000 kflop/s
1 Gflop/s = 1000 Mflop/s
1 Tflop/s = 1000 Gflop/s
1 Pflop/s = 1000 Tflop/s

3. Site context

3.1 Hosting Entity Overview

CINECA – funded in 1969 – is a not-for-profit Consortium, made up of 118 members: the Italian Ministry of Education, the Italian Ministry of Universities and Research, 70 Italian universities and 46 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 canter. 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 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 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 work for the managing, support, and exploitation of the HPC infrastructure, providing services to address computing research needs.

CINECA is appointed by the EuroHPC Joint Undertaking to execute the project in its data center facility.

3.2 Implementation

This procurement is a result of a strong collaboration between CINECA and EuroHPC Joint Undertaking.

The procured resources, integrated in the existing Leonardo infrastructure, will represent a new step forward to address the new challenges in the most diverse scientific fields. Paramount for the success of the project is:

- having the procured infrastructure operational as soon as possible to achieve a significant return of investment for the involved CAPEX.
- leverage the Leonardo infrastructure in order to improve both infrastructures.

The procured infrastructure will be installed beside Leonardo.

Moreover, the procured infrastructure will be covered by a maintenance service described in this document in Chapter 6 Maintenance and infrastructure availability.

3.2.1 Procedure

For this procurement, the EuroHPC Joint Undertaking elected to use a public open procedure. The goal of this document is to provide in the following all the requirements the Candidates will be requested to satisfy with their Offers.

3.2.2 Time schedule

The Procurer targets to accept the system by the 29st of August 2025 according to the timeline defined in the contract template part of the tendering package. To achieve this goal, the Procurer expect to:

- Complete the delivery of the system by April 2025
- Complete the installation process by July 2025
- Complete the acceptance process by August 2025.

3.3 CINECA facility description

The new procured equipment will be hosted in CINECA data centre located at *Big Data Technopole* - via Stalingrado 86, Bologna, Italy - beside Leonardo and in the same data hall. To set the logistic and data centre integration limits that the offers must comply, in the following the data centre specifications and architectural design are reported.

3.3.1 Data centre specifications

The floor space planimetry of the data centre is reported in Figure 2. For the data hall "HPC1" the following specifications apply:

Req.	Description	Category
3.3.1-1	*Dedicated data hall* The computing partitions of the procured infrastructure must be in installed the data hall "HPC1" where Leonardo is located. In terms of whitespace available, the data hall provides an area in the order 60 m^2.	DCS
3.3.1-2	*Raised floor details* The data hall is equipped with a raised floor with height of 100 cm.	DCS
3.3.1-3	*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
3.3.1-4	*Rack maximum height* The cable tray is at 235 cm of height.	DCS
3.3.1-5	*Room temperature* The hall is kept at the range temperature of 27-32°C.	DCS

3.3.1-6	*Room humidity* The offered infrastructure must comply with a relative humidity in the interval 20-60%.	DCS

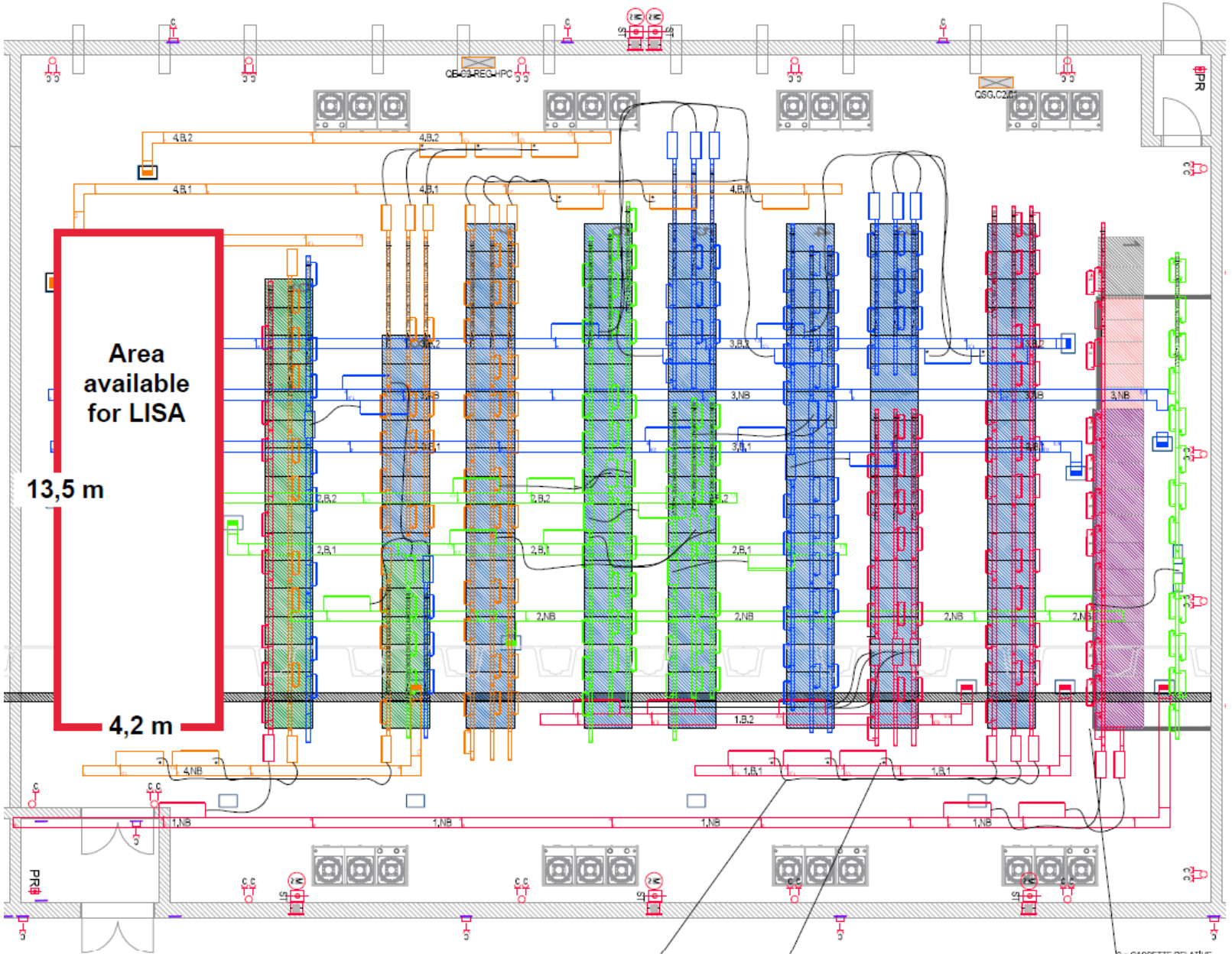

Figure 1: Area available for the LISA compute partition.

3.3.2 Electrical infrastructure

Req.	Description	Category
3.3.2-1	*Data hall power supply* The data hall HPC1 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 1.600 kW IT during the acceptance phase testing, and 1.200 kW IT in operating conditions.	DCS
3.3.2-2	*Electric load layout* The IT load installed in the data hall HPC1 is available through 2 rack rows. Every single rack will be connected to 3 dedicated busbars installed above the racks. Upon request of CINECA, the computing racks will be connected to the power supplies directly with cables, i.e., without the use of plugs and without compromising the warranty of the offer's components. The total power distributed to the racks cannot in any case exceed the limits defined in Req. 3.3.1-1.	DCS

3.3.3 Cooling infrastructure

The cooling of the procured infrastructure will rely on the cooling infrastructure already in use for Leonardo. Any change in the cooling water temperature set points will impact Leonardo operations – including its operating costs - and should be reduced to the minimum.

Req.	Description	Category
3.3.3-1	*Cooling infrastructure type* The cooling infrastructure of the data centre produces tempered water and chilled water. The tempered water is dedicated to the direct liquid cooling (DLC) of the compute nodes and the chilled water is used for dissipating what is not removed from DLC. The procured infrastructure needs to comply with this cooling infrastructure requirement and the limits reported in req. 3.3.4-2 and 3.3.4-3.	DCS
3.3.3-2	*Liquid cooling* Tempered water will be used for direct cooling of the system. The tempered water circuit is at 36°C inlet - as it is for Leonardo system. The tempered water circuit has a cooling capacity of 1.600 kWf dedicated for the procured infrastructure.	DCS
3.3.3-3	*Air cooling* The air-cooling system is based on 8 CRAH. Each CRAH machine has a cooling power of roughly 125 kWf. Considering two machines for redundancy, the cooling capacity made available for the the entire data hall is 750 kWf, of which 350 kWf dedicated for LISA. Optionally one more CRAH can be added to the room close to the LISA system.	DCS
3.3.3-4	*Flow rate* The maximum flow rate available to each rack row is 2100 l/min during acceptance tests and 1700 during operations.	DCS

3.3.4 Data Hall MEP layout

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

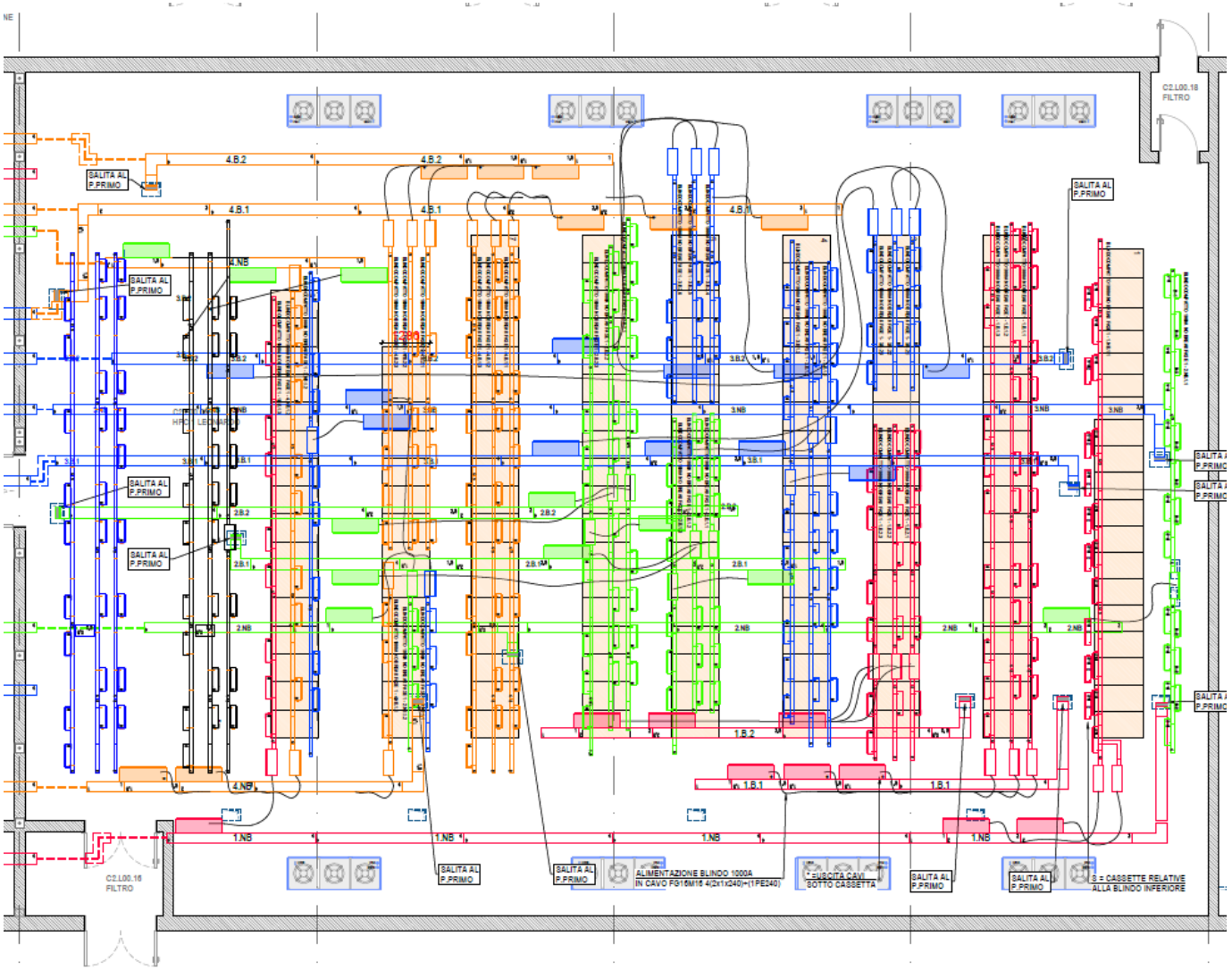

Figure 2: Power distribution layout of Data Hall 1. Each colour represents one of 4 electric branches.

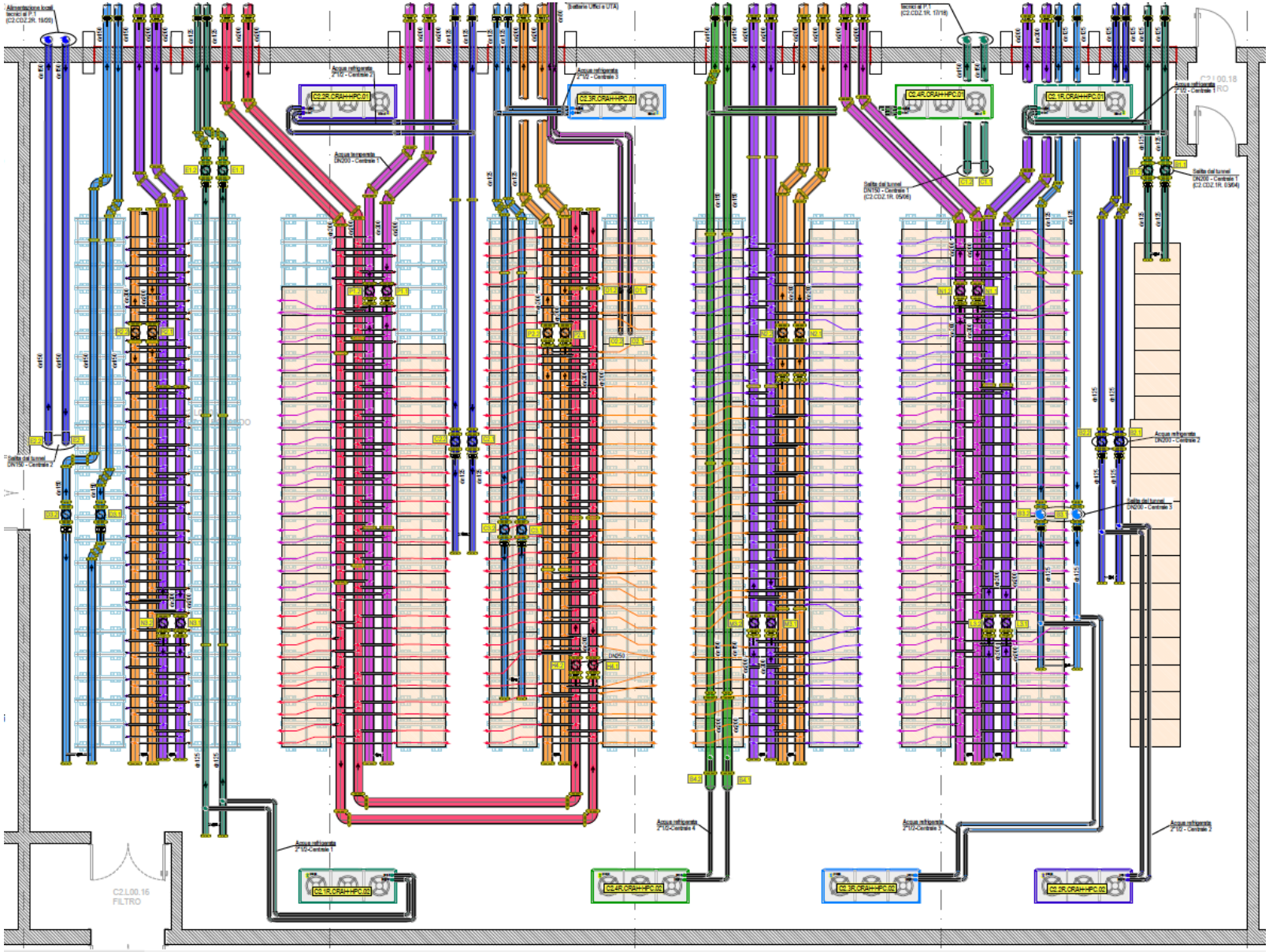

Figure 3: Cooling distribution layout of data hall 1.

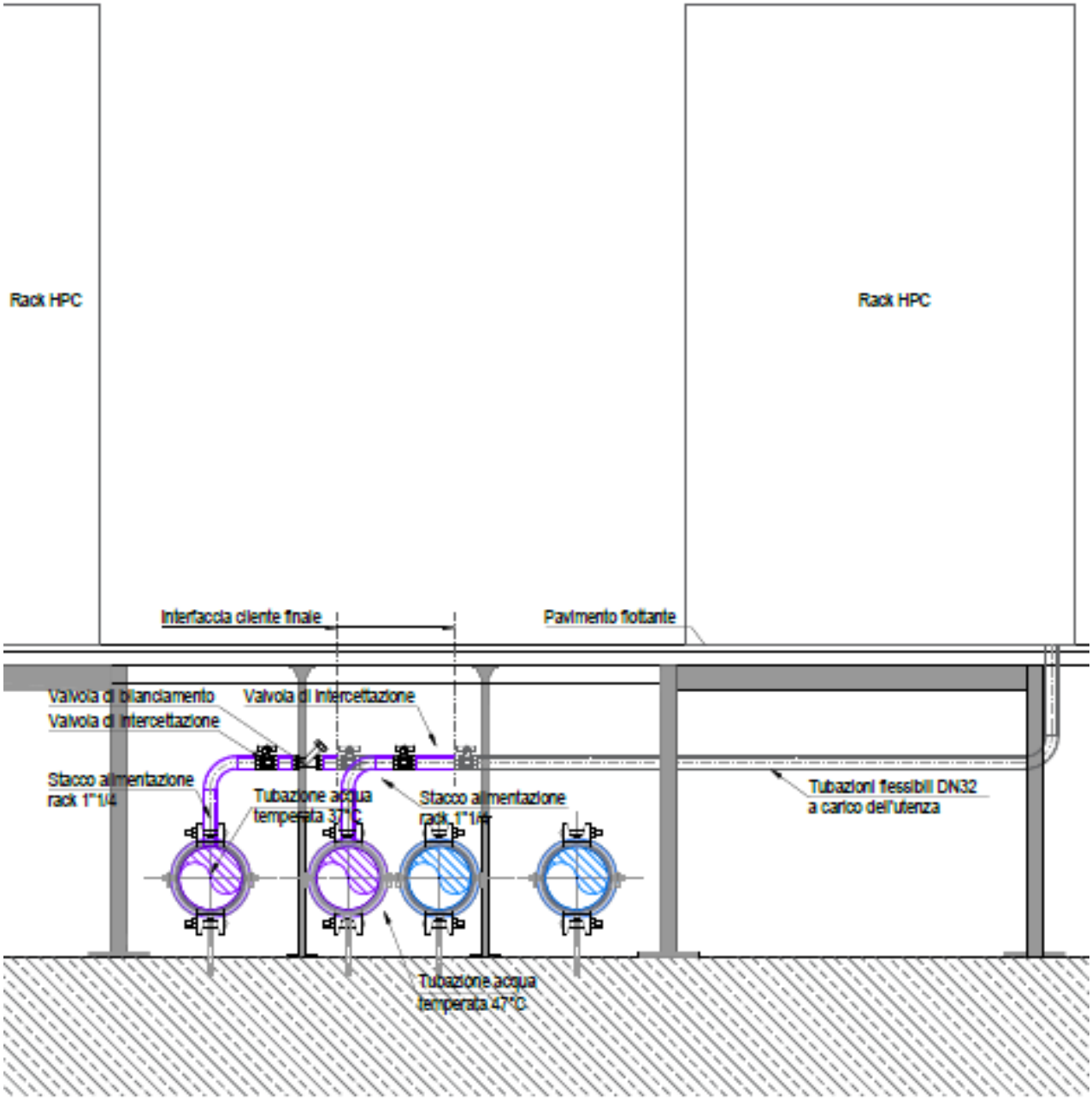

Figure 4: Vertical section of the Data Hall including rack and piping under the raised floor.

4. Technical Specifications

4.1 General requirements

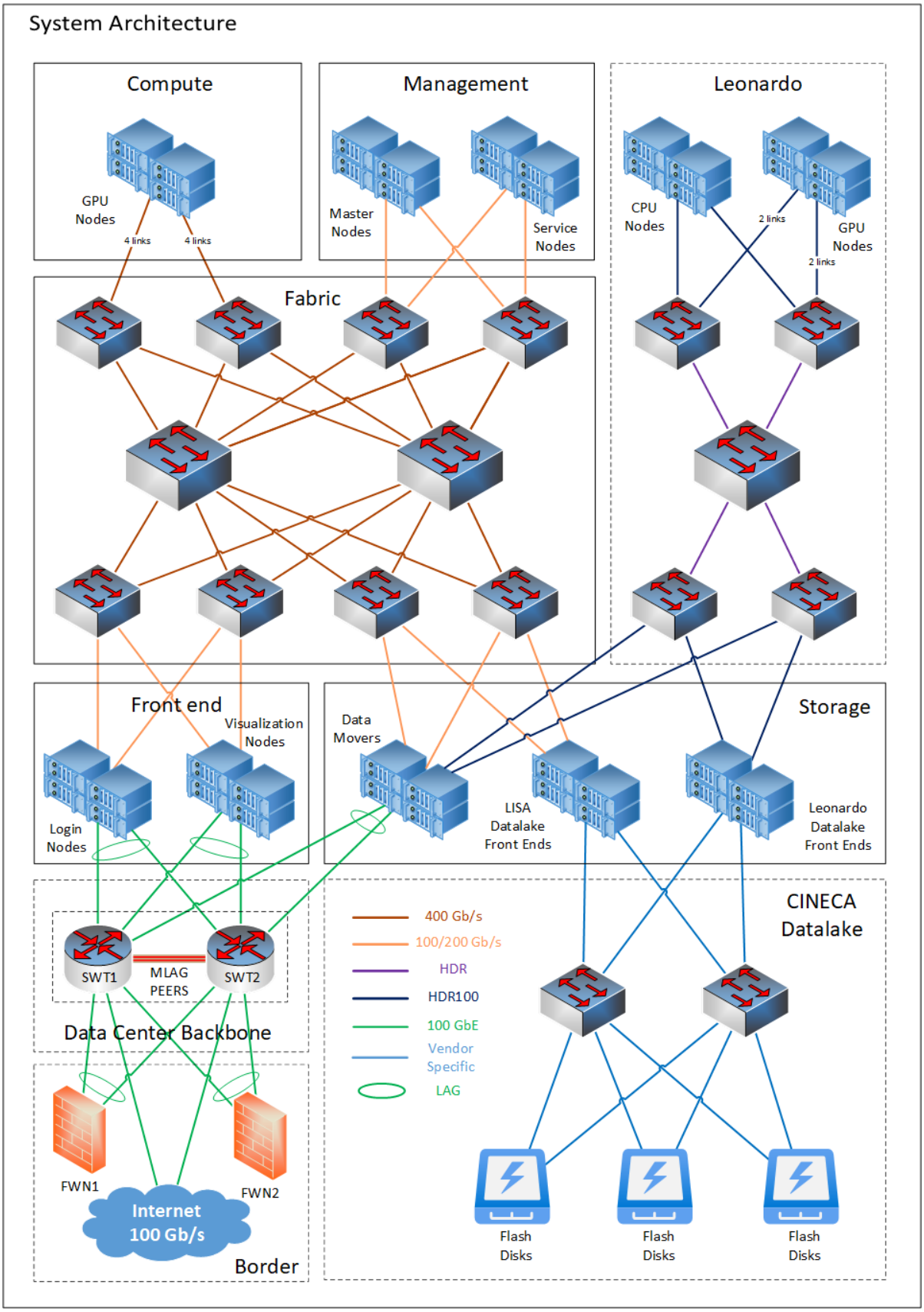

Figure 5: Reference design of the system architecture.

4.1.1 Functional aspects

Req.	Description	Category
4.1.1-1	*Integrated platform* The procured infrastructure is an integrated platform. All hardware and software components required to deliver service to users and manage the system must be included in the Offer.	MRQ
4.1.1-2	*Reboot time* Each partition must be fully rebooted in less than 60 minutes.	MRQ

4.1.1-3	*Common node features*
	
	The nodes must be equipped with:
	• Board Management Control (BMC) with the following features:
	  ○ Dedicated or shared Ethernet network port.
	  ○ Remote management protocols such as Java and HTML5 GUI.
	  ○ Virtual console & VMedia functionalities.
	  ○ System lockdown will prevent unintentional system configuration changes when system is running. When system lockdown is turned on, all system configuration changes including firmware updates will be prevented and user will be notified accordingly.
	  ○ Digitally signed firmware updates.
	  ○ Firmware rollback capabilities.
	  ○ Protection features for firmware updates of internal components.
	  ○ Secure default password functionality.
	  ○ LDAP authentication support.
	  ○ IP blocking functionality.
	  ○ Air flow management functionality.
	• Remote Monitoring and Alert Functionality:
	  ○ System capable of automatically sending an alert to the support, containing all the relevant information to diagnose the failure without any intervention from the System Administrators. Specifically, for RAM and disk drive components, when a pre-failure event is detected, the system must automatically send the alert.
		MRQ

4.1.1-4	*Health monitoring* The procured infrastructure must provide the capability to monitor the health parameters of each component via adequate software/hardware infrastructure. The monitoring software infrastructure should expose open-source API/frameworks in order to be integrated with open-source tools. All hardware faults of the components that can affect the performance and stability of the nodes and devices of the system must be reported.	MRQ
4.1.1-5	*Node power and energy measurement* The procured infrastructure must provide node power and energy measurements with a minimum of 95% accuracy and low impact on	MRQ

	performance. The details of the implementation will be considered in evaluating the Offers that provide this capability.	
4.1.1-6	*Monitoring APIs* The monitoring and management systems of the procured infrastructure must provide APIs enabling the integration with third-party monitoring and management frameworks. APIs must provide information on life status of all the components of the infrastructure in a timely fashion to be alerted on incurring faults within 300 seconds from their occurrence.	MRQ

4.2 Interconnects

4.2.1 Fabric

Req.	Description	Category
4.2.1-1	*General requirements* The procured infrastructure must provide a low-latency high bandwidth fabric used to interconnect the nodes. The fabric must implement the following features: • Must be based on at least 400 Gb/s. • The minimum bandwidth for each access link must be at least 100 Gb/s with full bi-directional bandwidth per port. • Support to access ports at 200Gb/s and 400 Gb/s. • Support RDMA communications. • The average latency of MPI Point-to-Point communication must be less than 3 microseconds. • Provide optimization for MPI communications. • Support Fat-tree and/or Dragonfly(+) topology. The fabric is also represented in the top part of Figure 7.	MRQ
4.2.1-2	*Bandwidth* A fabric with full bisection bandwidth is required.	MRQ
4.2.1-3	*Topology* A full-fat tree (non-blocking) topology is required.	MRQ
4.2.1-4	*Participants* All nodes must be connected to the fabric as shown in Figure 3.	MRQ
4.2.1-5	*Monitoring and managing capabilities.* The following capabilities must be provided: • The fabric must provide managing mechanism and for near-real time collection of performance and health information. • Each switch of the fabric must provide an out-of-band management port based on 1 GbT.	MRQ
4.2.1-6	*Advanced fabric features* The fabric should support the following mechanisms:	MRQ

	• *In-network computing*: for collective offloads. • *Avoid congestion*: through adaptive routing mechanisms. • *Self-healing*: with re-routing based on auto-discovery changes of the topology.	
4.2.1-7	*Power redundancy* Each compute fabric switch must be equipped with redundant & hot-swappable power supplies.	MRQ

4.2.2 Management Network

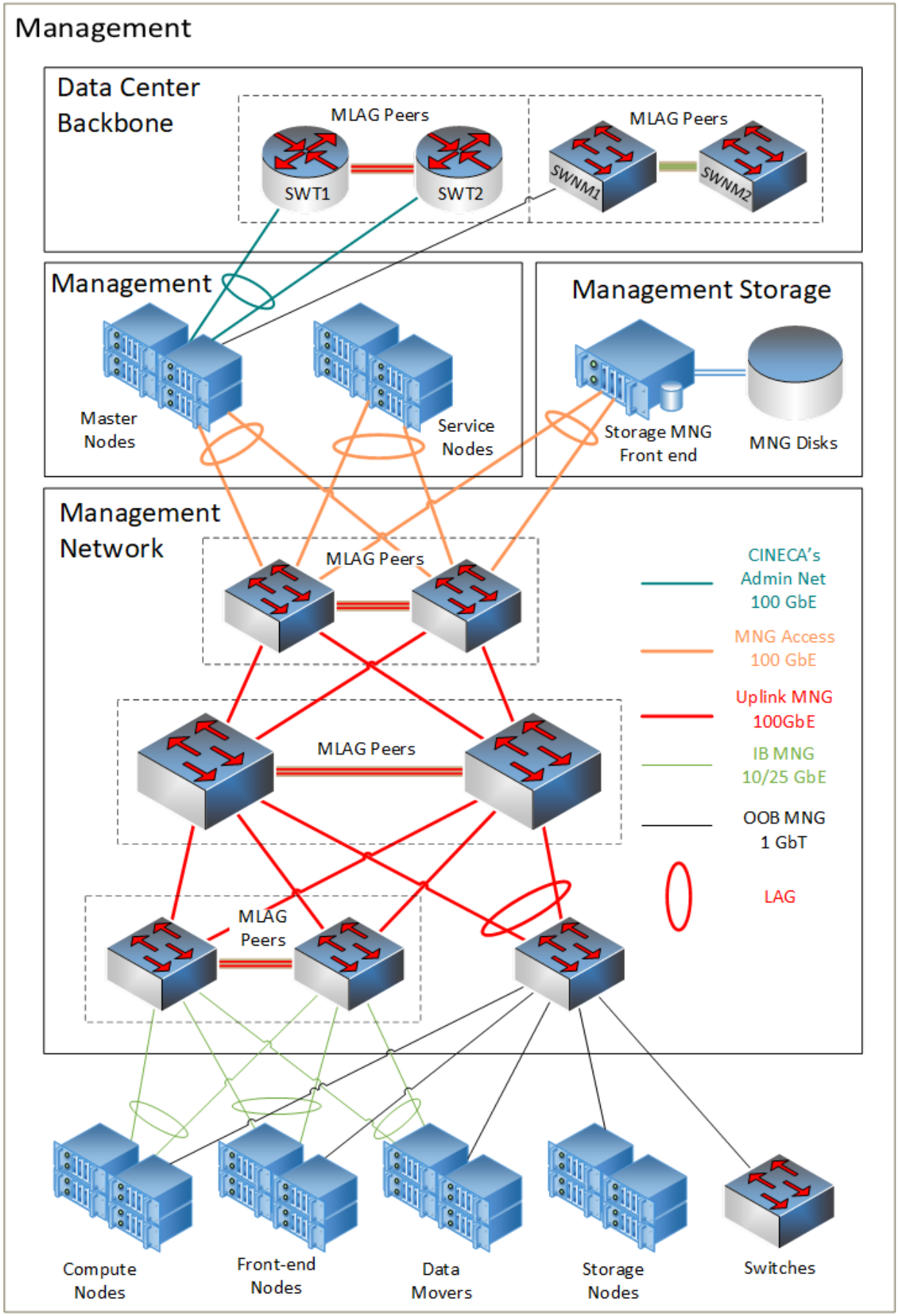

Figure 6: Reference design of the management network.

Req.	Description	Category
4.2.2-1	*Management network* The system must provide a physically dedicated Ethernet network for management purposes with the following characteristics: • Spine-leaf layer 2 based on MLAG topology is preferred. • The oversubscription of the topology must not exceed 4:1. • The network topology must be redundant both at inter-switch links and aggregation switches. Failures of single switch/link must not affect the network stability except for access switches. • The management network must support two sub networks at logical link layer (VLAN): *In-Band (IB) and Out-Of-Band (OOB) management (MNG) network*. The management network is represented in Figure 4.	MRQ
4.2.2-2	*Network Participants* • I*n-Band (IB) management network*: used for managing and deploying compute nodes and data movers for operational services: bare metal/OS installation, OS monitoring and metering, etc. • *Out-Of-Band (OOB) management network*: used for managing Board Management Controller (BMC) of all system's equipment (storage, networks, chassis, etc.).	MRQ
4.2.2-3	*Monitoring and managing capabilities* The following capabilities must be provided: • The management network must provide methods for management and for near-real time collection of performance and health information (e.g., sFlow). • Each switch of the management network must provide an OOB MNG port based on 1GbT connected to the OOB MNG network.	MRQ
4.2.2-4	*Aggregation switches* • Switches at aggregation level must provide ports at least 100 GbE to be connected to the access switches and to the management nodes. The number of aggregation switches and the number of ports will be defined by the Candidate in order to comply with the req. 4.2.3-1. • Aggregation switches must support static and dynamic routing (Layer 3) and relative protocols (e.g., static routes, OSPF, BGP, MP-BGP, OSPFv3, etc.).	MRQ
4.2.2-5	*Access switches* • Must provide access ports at 1 GbT for OOB MNG ports. • Must provide access ports at 10/25 GbE for IB MNG ports. • Must provide access ports at 100 GbE for the Management Nodes and storage. • Must provide uplink ports at 100 GbE.	MRQ
4.2.2-6	*Network capabilities* The switches of the management network must support:	MRQ

	• Full support to IPv4 and IPv6. • MLAG, LAG, LACP, VLAN protocols and most common IEEE 802.X network protocols.	
4.2.2-7	*Network performance* • Performance of these networks will allow for the full reconfiguration of the OS (without re-installation) of all nodes in less than 2 minutes. • Performance of these networks will allow for the (re-) installation of the OS of all CN nodes in less than 3 hours. • Performance of these networks will allow for cold reboot of all CN nodes in less than 60 minutes measured from the shut-down of the first node to the boot of the last non-faulted node. • Performance of these network will allow to collect all metrics and sensors of the management board of all compute nodes and network devices with open tools (i.e., IPMI tool, Redfish, Confluent, SNMP) in less than 20 seconds using as many parallel sessions as the monitoring infrastructure can use.	MRQ
4.2.2-8	*Power redundancy* Each switch in the management network should be equipped with redundant & hot-swappable power supplies.	TRQ

4.3 Compute Partition

Req.	Description	Category
4.3-1	*Partition performance* The partition must feature at least 165 nodes.	MRQ
4.3-2	*Common node requirements* The nodes must implement the requirements provided in Req. 4.1.1-3.	MRQ
4.3-3	*CPU Technology* The CPU must be based on x86_64 architecture.	MRQ
4.3-4	*GPU technology* The GPUs must support the following characteristics: • State-of-the-art GPUs with support to AI training in particular for LLMs. • Provide memory sharing with all the other GPUs installed in the same node. • GPU's HBM memory must provide at least 80 GBytes.	MRQ
4.3-5	*Node configuration* Each node must be equipped with 2 CPUs and 8 GPUs (8-way GPU configuration).	MRQ
4.3-6	*DRAM memory* • The node's memory must be at least 1 TByte of DDR5 memory and greater equal to the sum of all GPU memories installed in the node.	MRQ

	• The nodes must be configured to saturate all DDR memory channels of the CPUs (or in an optimal configuration to saturate the available memory bandwidth).	

4.3-7	*Network requirements*
	
	The nodes must be equipped with:
	• 1 NIC per GPU connected to the fabric with at least 400 Gb/s (≥3.2 Tb/s aggregated). No more than 2 NICs must be connected on the same switch as shown in Figure 6.
	• 1 NIC with 2 Ethernet ports connected to the IB MNG network at 10 Gb/s or 25 Gb/s. The network ports must be connected to different switches, and they must support:
	  o Pre-execution environment (PXE) boot.
	  o Remote boot over Ethernet.
	• 1 NIC with 1 Ethernet port dedicated to the BMC and connected to the OOB MNG network at 1Gb/s.
	
	OOB and IB MNG NICs can also share the same physical port.
		MRQ

4.3-8	*Node local storage* The nodes must be equipped with: • x2 SSD drives in RAID1 configuration with net space available of ≥ 0.8 TByte for the OS. • x8 SSD drives with a net space available for each drive of at least ≥ 3 TBytes used for a scratchpad storage. The drive array should support multiple RAID configurations such as RAID0/1/5/6/10 and global hot spare drives.	MRQ
4.3-9	*Fast distributed scratchpad* This partition should be supported with a dedicated fast distributed scratchpad built on top of the SSD drives equipped within the nodes and described in req. 4.3-8. This scratchpad should be accessible to all the compute, login, visualization nodes and data movers. It should provide the following target performance: • At least 2.5 PBytes of net space available. • At least a total read throughput of 9 TB/s. • At least a total write throughput of 2 TB/s. • At least 300M IOPS (4k random reads) for parallel file system (or NFS) reads. • At least 90M IOPS (4k random writes) for parallel file system (or NFS) writes.	TRQ

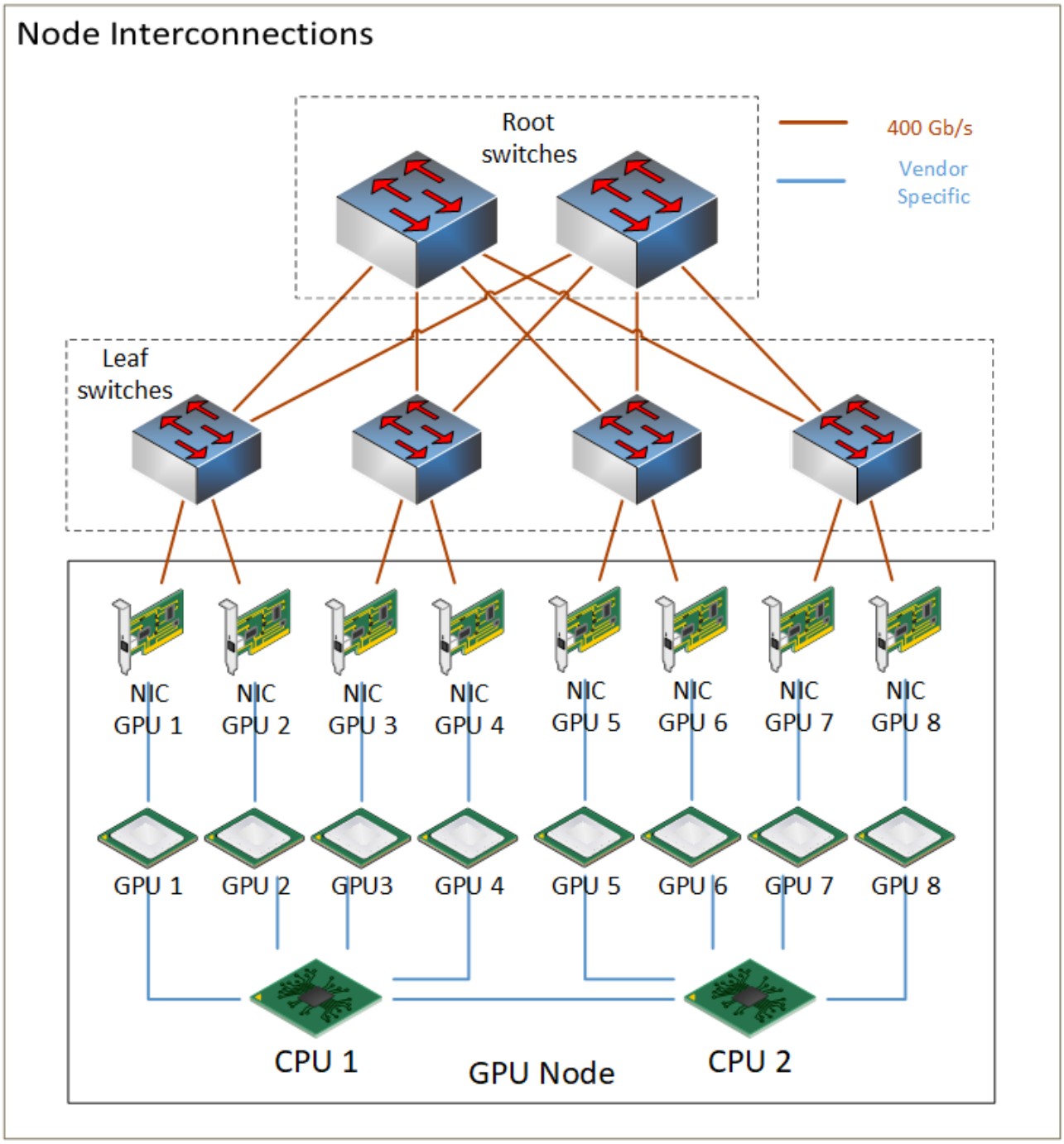

Figure 7: Reference design of the GPU and Fabric interconnection. The internal node connection will be defined by the Candidate depending on the specific GPU technology.

4.4 Management partition

The Management partition includes two sub partitions and a shared storage:

- *Service partition*: A set of nodes dedicated to host all the general critical services (e.g., workload schedulers, system monitor, etc.).
- *Master partition*: A set of nodes dedicated for the whole cluster management (bare metal provisioning, internal networks management, etc.).
- *Management storage*: a shared storage infrastructure used from the Management Nodes to archive and collect information from the System.

Req.	Description	Category
4.4-1	*General requirements* The management partition is used to host all system services and management tools of the System. The size of the management partition must be sufficient to support the operation of the system. This partition will feature no less than 8 nodes. Candidates are invited to employ virtualization techniques to reduce the size of the service partition while complying with the above minimum.	MRQ
4.4-2	*Common node requirements* The nodes must implement the functional requirements provided in Req 4.1.1-3.	MRQ
4.4-3	*Node configuration* The nodes must be equipped with two CPUs based on x86_64 architecture.	MRQ
4.4-4	*Memory configuration* All Management Nodes must feature a total of at least 256 GBytes of DDR5 memory.	MRQ
4.4-5	*Management storage* Management Nodes must have a shared storage with at least 200 TBytes to contain: • All the management software. • All the management databases and an historical daily (differential) backup of these databases for a year. • The aggregated system logs of all the node partitions for at least one year. • The aggregated audit logs of Compute, Front end, and Management Nodes for at least two years. • The performance and functional metrics collected from all nodes and equipment for at least two years. • The shared storage must be connected to the management network and must be possible to be mounted from all Management Nodes. • The management storage must be physically separated from the storage of the other partitions and must be shared among all the management nodes and seamlessly available to all the services. • The storage configuration must be resilient to the failure of at least two independent basic blocks (i.e., storage nodes, controllers, or disk chassis).	MRQ
4.4-6	*Networking configuration* Each Management Node must be equipped with: • 1 NIC with 2 ports connected to the fabric with at least 200 Gb/s (400 Gb/s aggregated). • 1 NIC with 2 Ethernet ports connected to the CINECA's Admin Network with 100 Gb/s (200 Gb/s aggregated).	MRQ

	• 1 NIC with 2 Ethernet ports connected to the IB MNG network with 100 Gb/s (200 Gb/s aggregated). The network ports must be connected to different switches and they must support:
	  ○ Pre-execution environment (PXE) boot.
	  ○ Remote boot over Ethernet.
	• 1 NIC with 1 Ethernet port dedicated to the BMC and connected to the CINECA's OOB MNG network at 1 GbT.
	
	OOB and IB MNG NICs can also share the same physical port.	

4.4-7	*Connection with* CINECA's HPC Backbone The Management Nodes must be connected with the CINECA's HPC Backbone Network, which is based on two NVIDIA Spectrum SN4700[1] (represented in Figure 4 as *SWT1* and *SWT2*) with Cumulus Linux. The cabling length required for the connection is between up to 100 meters depending on the final racks' layout. All cables, transceivers, and 100/400 GbE splitters for both Management Nodes and CINECA's network equipment (NVIDIA Spectrum SN4700) must be provided in the Offer. See Figure 4 for the connection's details of the Management nodes.	MRQ
4.4-8	*Performance* The number of the Service Node must guarantee the requested performance levels for installation and reboot as well as effectiveness in collecting, storing, and processing all the metrics and logs. An excellent performance level must be guaranteed also during queries to collected data and all the management, troubleshooting, accounting, and security assessment activities.	MRQ
4.4-9	*High Availability* The service partition must include the required hardware components to configure all important system services in high availability. The service nodes must be configured in "cluster" mode, meaning that they should guarantee fault tolerance in terms of hardware and software services. All these functionalities must be available even in case of single or double node's fault with adequate levels of efficiency. Besides, a workload running on the Compute, Visualization and Login Nodes must be able to continue working without significant interruption. Any performance impact will be described in the Offer.	MRQ
4.4-10	*Health and Consistency Checks* The Candidate will provide tools to check the health and validate configuration of hardware and software components that can be integrated with the workload manager to ensure that only fully functional components are utilized for jobs. Where applicable, auto recovery actions will be performed and logged.	TRQ
4.4-11	*Rolling Updates* The system will provide "rolling update" mechanisms that allow reliable software updates and selected maintenance operations to be performed with	MRQ

[1] https://www.nvidia.com/content/dam/en-zz/Solutions/networking/br-sn4000-series.pdf

	minimal accumulated downtime. In full system maintenances, the idle time of Nodes incrementally grow prior to the start of the maintenance as running jobs finish and no new jobs start due to the pending maintenance reservation. The requested feature will significantly reduce this maintenance overhead.	
4.4-12	*Cluster Management Software* The Candidate will provide an integrated software solution for the management of all cluster resources, the provisioning of nodes and basic hardware and operating system monitoring. The software will offer support for the (out-of-band) management of all hardware components and node provisioning. The software will enable the automation of all the fundamental system management activities: • Installation of the OS on the nodes. • Reconfiguration of the OS of the nodes (and possibly of all apparatus). • Collection of nodes diagnostic information (and possibly of all apparatus). • Update of the firmware nodes. This software is typically in execution on the Service nodes, must feature a redundant configuration mechanism, and preferably be open source and belonging to the OpenHPC initiative.	MRQ
4.4-13	*Basic Hardware Monitoring* The cluster management software will provide out-of-band and/or in-band monitoring of hardware events (e.g., system event log and machine check exceptions, if applicable). The events will be collected and stored at a central location.	MRQ

4.5 Front-end partition

The Front-end partition includes two sub partitions:

- *Login partition*: for external system access, compilation and data management activities, job submission as well interactive pre-/post-processing workloads.
- *Visualization partition*: to enables visualization of simulation results during and after the execution of jobs. Visualization Nodes may be operated in batch mode or as externally accessible interactive nodes.

4.5.1 Login partition

Req.	Description	Category
4.5.1-1	*Partition size* The login partition must feature at least 16 nodes.	MRQ
4.5.1-2	*Common node requirements* The nodes must implement the functional requirements provided in Req. 4.1.1-3.	MRQ
4.5.1-3	*Node configuration* The Login Nodes must feature:	MRQ

	• the same CPUs of the compute nodes. • at least one GPU with the same technology of the compute nodes.	
4.5.1-4	*Memory configuration* The Login Nodes must feature a total of at least 512 GBytes of DDR5 memory.	MRQ

4.5.1-5	*Network requirements*
	
	All Login Nodes must be equipped with:
	• 1 NIC with 2 ports connected to the fabric with at least 200 Gb/s (400 Gb/s aggregated). The network ports must be connected to different switches.
	• 1 NIC with 2 Ethernet ports connected to the Data Center Backbone network with at least 100 Gb/s per port (200 Gb/s aggregated). The network ports must be connected to different switches.
	• 1 NIC with 2 Ethernet ports connected to the IB MNG network at 10 Gb/s or 25 Gb/s. The network ports must be connected to different switches and must support:
	  ○ Pre-execution environment (PXE) boot.
	  ○ Remote boot over Ethernet.
	• 1 NIC with 1 Ethernet port dedicated to the BMC and connected to the OOB MNG network at 1Gb/s.
	
	OOB and IB MNG NICs can also share the same physical port.
		MRQ

4.5.1-6	*Local storage* The nodes must be equipped with x2 SSD drives in RAID1 configuration with net space available of ≥ 7 TBytes for the OS.	MRQ
4.5.1-7	*Software installation* Login Nodes must allow the installation of all user software and applications that need to be run on the system.	MRQ

4.5.2 Visualization partition

Req.	Description	Category
4.5.2-1	*Partition size* The visualization partition must be composed to at least 4 nodes.	MRQ
4.5.2-2	*Common node requirements* The nodes must implement the functional requirements provided in Req. 4.1.1-3.	MRQ
4.5.2-3	*Node configuration* The nodes must be equipped with: • Two state-of-the-art CPUs, binary compatible with the CPU equipping the compute nodes. • at least two high-end graphical cards supporting 3D acceleration through OpenGL graphics.	MRQ
4.5.2-4	*Memory configuration*	MRQ

	The Visualization Nodes must feature a total of at least 512 GBytes of DDR5 memory.	

4.5.2-5	*Network requirements*
	
	All Login Nodes must be equipped with:
	• 1 NIC with 2 ports connected to the compute fabric with at least 200 Gb/s (400 Gb/s aggregated). The network ports must be connected to different switches.
	• 1 NIC with 2 Ethernet ports connected to the Data Center Backbone network with at least 100 Gb/s per port (200 Gb/s aggregated). The network ports must be connected to different switches.
	• 1 NIC with 2 Ethernet ports connected to the IB MNG network at 10 Gb/s or 25 Gb/s. The network ports must be connected to different switches, and they must support:
	  ○ Pre-execution environment (PXE) boot.
	  ○ Remote boot over Ethernet.
	• 1 NIC with 1 Ethernet port dedicated to the BMC and connected to the OOB MNG network at 1 Gb/s.
	
	OOB and IB MNG NICs can also share the same physical port.
		MRQ

4.5.2-6	*Local storage* The nodes must be equipped with x2 SSD drives in RAID1 configuration with net space available of ≥ 7 TBytes for the OS.	MRQ

4.6 Storage infrastructure

4.6.1 Data Movers

Req.	Description	Category
4.6.1-1	*Partition size* The Data Mover partition must be composed to at least 3 nodes.	MRQ
4.6.1-2	*Common node requirements* The nodes must implement the functional requirements provided in Req. 4.1.1-3.	MRQ
4.6.1-3	*Node configuration* The nodes must be equipped with two CPUs based on x86_64 architecture.	MRQ
4.6.1-4	*Memory configuration* The nodes must feature a total of at least 256 GBytes of DDR5 memory.	MRQ
4.6.1-5	*Network requirements* All Login Nodes must be equipped with: • 1 NIC with 2 ports connected to the fabric with at least 200 Gb/s (400 Gb/s aggregated). The network ports must be connected to different switches. • 1 NIC with 2 HDR100 ports connected to the Leonardo fabric. The network ports must be connected to different switches.	MRQ

	• 1 NIC with 2 Ethernet ports connected to the Data Center Backbone network with at least 100 Gb/s per port (200 Gb/s aggregated). The network ports must be connected to different switches.
	• 1 NIC with 2 Ethernet ports connected to the IB MNG network at 10 Gb/s or 25 Gb/s. The network ports must be connected to different switches, and they must support:
	  ○ Pre-execution environment (PXE) boot.
	  ○ Remote boot over Ethernet.
	• 1 NIC with 1 Ethernet port dedicated to the BMC and connected to the OOB MNG network at 1Gb/s.
	
	OOB and IB MNG NICs can also share the same physical port.
	Connection details of the Data Movers are represented in Figure 3.	

4.6.1-6	*Local storage* The nodes must be equipped with x2 SSD drives in RAID1 configuration with net space available of ≥ 7 TBytes for the OS and local files.	MRQ

4.6.2 Datalake front ends

The System will not be equipped with its own dedicated global storage, but it will rely on the CINECA's datalake infrastructure. The datalake will be also connected to Leonardo therefore providing a shared storage for both systems. For this reason, datalake front-end nodes for LISA and for Leonardo are required. The CINECA's datalake is based on VAST Data technology[2]. Figure 4 shows the interconnection details of CINECA's datalake and Figure 7 shows the location of CINECA's datalake and the Leonardo's L2 switches that will be used to uplink the datalake.

4.6.2.1 LISA front ends

Req.	Description	Category
4.6.2.1-1	*Partition size* The LISA's datalake connections must be composed of at least 28 storage front ends. Compatibility with the CINECA's datalake deployment is required.	MRQ
4.6.2.1-2	*Location of datalake front ends* The storage front ends must be installed in the CINECA's datalake location as shown in Figure 7. For this reason, 2 standard 19" 42U racks must be provided (fully equipped with PDUs) and installed close to the CINECA's datalake racks. Each rack must contain half of the storage front ends for LISA and the L1 switches of the LISA's fabric to connect the storage front ends with the compute nodes.	MRQ

[2] https://www.vastdata.com/whitepaper.pdf

Figure 8: CINECA's datalake and Leonardo's interconnection location.

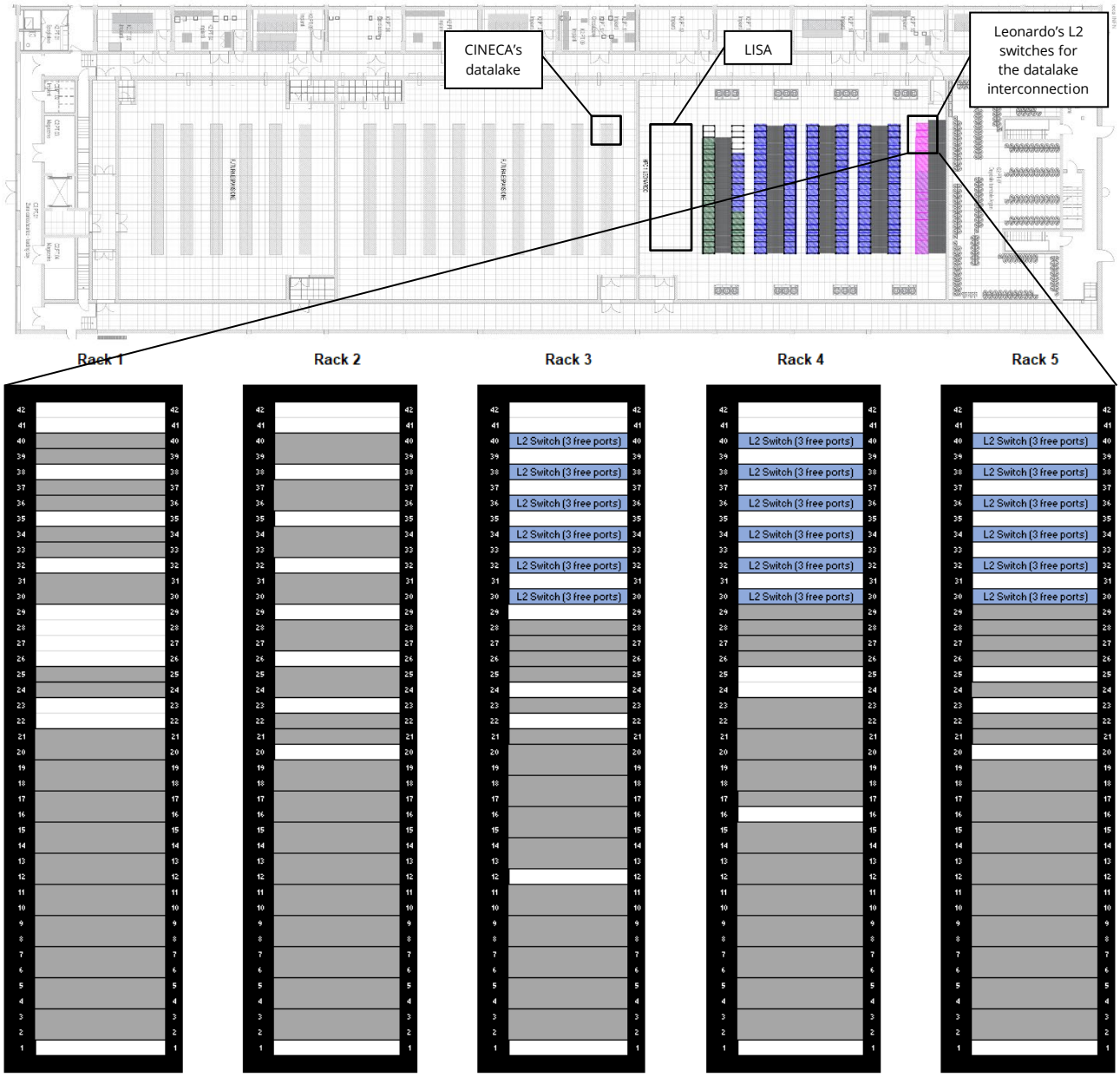

Figure 9: Rack layout of the I/O L2 switches in the Leonardo's fabric.

4.6.2.2 Leonardo front ends

The Leonardo's fabric is an HDR InfiniBand network based on Nvidia Quantum QM8700[3]. The topology is based on a Dragonfly+ with two-tier organisation. Nodes are built into fat-tree groups, which are then connected in an all-to-all fashion as shown in Figure 9-a. Each group is composed of 18 L2 switches that connect one group to the others. Each group contains 18 L1 switches connected with 18 uplinks to the L2 switches (one uplink per L2 switch) and with 22 downlinks to the nodes. There are four types of Dragonfly+ groups: CPU, GPU, Hybrid (mix of CPU and GPU), and I/O. The I/O group contains the storage, logins, service nodes and it is represented in Figure 9-b. Differently from the others, the I/O group is composed of only 15 L1 switches, for this reason is possible to extend the I/O L1 switches (up to 3) in order to accommodate the Leonardo's datalake.

[3] https://network.nvidia.com/files/doc-2020/pb-qm8700.pdf

Req.	Description	Category
4.6.2.2-1	*Partition size* The Leonardo's datalake connections must be composed of 20 storage front ends. Compatibility with the CINECA's datalake deployment is required.	MRQ
4.6.2.2-2	*Leonardo's switches* The fabric of Leonardo must be extended with 2 Nvidia Quantum QM8700 or equivalent InfiniBand switches (compatibility with the Leonardo's fabric is required) to host the storage front ends.	MRQ
4.6.2.2-3	*Accessories for Leonardo's switches* In order to connect the switches of req. 4.6.2.2-2 to the L2 switches of the Leonardo's I/O group (Figure 8), are needed 36 InfiniBand HDR AOC cables with a length up to 100m (e.g. Nvidia MFS1S00-H1XXXE[4] or compatible) or in alternative: • 72 HDR QSFP56 SR4 MMF transceivers certified (or compatible) for Nvidia Quantum QM8700 (e.g. Nvidia MMA1T00-HS[5]). • 36 parallel multi-mode optical cable with standard MPO-12 UPC connectors with a length up to 100m certified (or compatible) for the above transceivers. The switches must be connected with 1 HDR cable to every L2 switches hosted in the racks 3, 4, and 5 (see Figure 8) in order to respect the Dragonfly+ topology.	MRQ
4.6.2.2-4	*Location of storage front ends* The storage front ends must be installed in the CINECA's datalake location as shown in Figure 7. For this reason, 2 standard 19" 42U racks must be provided (fully equipped with PDUs) and installed close to the CINECA's datalake racks. Each rack must contain half of the storage front ends for Leonardo and one L1 switch of the Leonardo's fabric to connect the storage front ends with the compute nodes.	MRQ

[4] https://network.nvidia.com/related-docs/prod_cables/PB_MFS1S00-HxxxE_200Gbps_QSFP56_AOC.pdf
[5] https://network.nvidia.com/pdf/prod_cables/PB_MMA1T00-HS_HDR_QSFP56_MMF_Transceiver.pdf

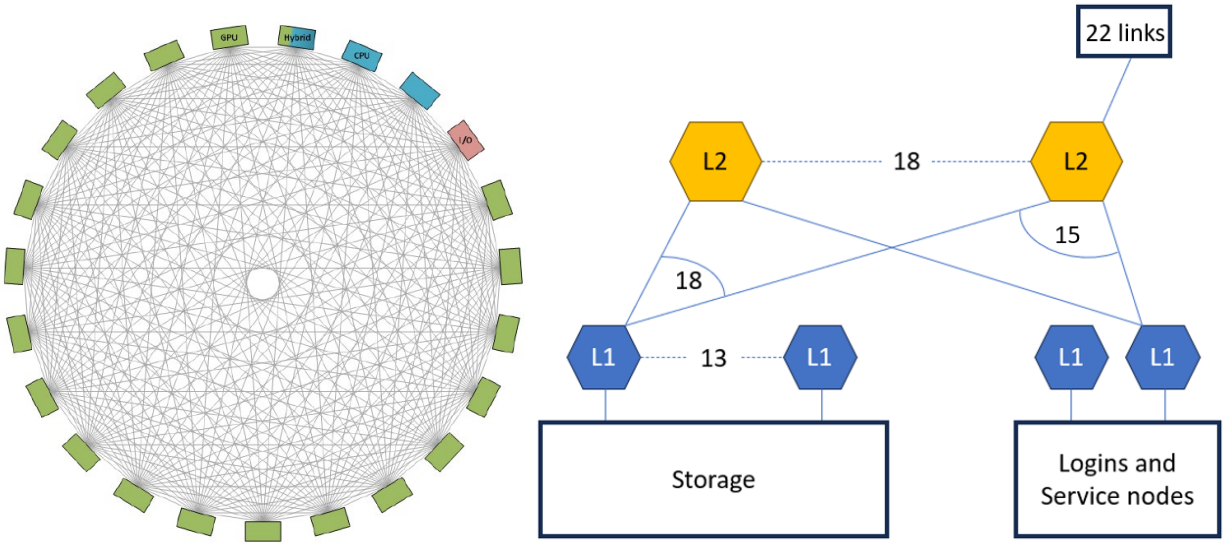

Figure 10: Of the left figure (a) is represented the Dragonfly+ topology of the Leonardo's fabric. On the right figure (b) is shown the organization of the I/O island.

4.7 System software and monitoring

Req.	Description	Category
4.7-1	*Compute Node Operating System* The operating system for all nodes will be RHEL with 64-bit kernel version 5.14 or higher, and supporting remote management, network boot and system image delivery. The same version of Linux should be present on all the node of the compute and front-end partitions.	MRQ
4.7-2	*Service Node Operating System* The operating system for the Service Nodes will be RHEL with 64-bit kernel version 5.14 or higher. The operating system on system critical components must be offered with full support.	MRQ
4.7-3	*Front end node Operating System* The operating system for all nodes will be RHEL with 64-bit kernel version 5.14 or higher, and supporting remote management, network boot and system image delivery. It will be allowed to install security patches soon after their release independently from the constraints of the CN (I.e. GPU or FS drivers/software)	MRQ
4.7-4	*User Management with LDAP* The design of the procured infrastructure will enable integration in CINECA's LDAP-based directory service for user management in such a way that no single-point of failures exist.	MRQ
4.7-5	*Container workload support* The offered solution must enable the remote building and execution of containerized applications, to better support AI workloads, providing all the requested features to enable this service (e.g. system software, container repository storage, licensing, and so on). The system will provide a mechanism	MRQ

	to build and modify containers from the front-end partition. At least one common container format will be supported, such as the Open Container Initiative (https://www.opencontainers.org/) v1.1.0 (or later) or Image Specification used by Singularity's image format (https://sylabs.io/). The solution will provide role-based access control to enable compliancy to site security policies.	
4.7-6	*Support for recent programming environment standards.* All offered MPI implementations and compiler suites will support recent versions of the applicable standards: • MPI version 4.0 or newer, • C ISO/IEC 9899:2011 or newer, • C++ ISO/IEC 14882:2014 or newer, • Fortran ISO/IEC 1539-1:2010 (aka Fortran 2008) or newer, • OpenMP 4.5 or newer Full stack software programming paradigm for accelerators, including at least C, C++ and Fortran front-end.	MRQ
4.7-7	*Lightweight Performance Profiling* The procured infrastructure will provide lightweight performance profiling capabilities that can be activated by the users on a job base. Data at process, job and node level will be made available (utilizing scalable accumulation methods). Data retention times for the mentioned granularity levels may differ. The technology will have minimal impact on application performance (less than 5% performance drop) and in particular not affect scalability of large jobs. A basic set of data must be gathered irrespectively of the user application, i.e., without requiring the users to link against specific libraries. At least the following system components will be covered: • CPU utilization (load avg.), IPC, Instruction mix information; • memory footprint, cache utilization/hits/miss, TLB hits/miss, load/store ops, memory interface utilization; ○ For systems featuring multiple memory types and a deep(er) memory hierarchy, information will be gathered for all types and tiers. • I/O subsystem: number of reads/writes, read/write bandwidth; • Network: number of packets/reads/writes (RDMA), packet/segment length. • Accelerator: utilization. • MPI/communication libraries: Number of calls, time spent. It will be possible to flexibly extend the gathered observables (potentially with additional/higher overhead).	TRQ
4.7-8	*Performance report generation* The Candidate will provide tools and/or an API to create job performance reports for users based on the collected data. Ideally, the focus of the report will be controllable in terms of the level of detail as well as the considered system components. The level of detail and numbers of levels must be dynamically adjustable by administrators and users. An API for accessing the	TRQ

	reports in a machine-readable format will be available (e.g., for the integration with the workload manager or external web portals). An integration with the workload manager, allowing appending reports to job output and e-mail based job notifications, is desirable. This integration requirement does not apply to offers that do not include the workload manager.	
4.7-9	*Optimized numerical libraries* The Candidate will provide highly optimized libraries providing API compatible replacements for BLAS, LAPACK and ScaLAPACK routines. The Proposal will include an optimized fast Fourier transform (FFT) library.	MRQ
4.7-10	*Parallel debugger* An adequate parallel debugger and profiler must be included in the software package to allow debugging of parallel application. It must be licensed for at least 10 CN.	MRQ
4.7-11	*Hardware counters* The hardware counters of the CPU/GPU should be mature and accessible e.g., by a tool like LIKWID. In addition, a software tool must be provided to measure and collect automatically the performance of the users' applications running as batch job. The performance measurement should be based on the metrics of the performance counters.	MRQ

5. Benchmarks

5.1 Introduction

This section describes the context and the main objectives of the benchmark procedure to evaluate the performance of solutions proposed by the Candidate.

The chapter is organized as follows:

- In Section 5.2 the benchmark framework is described, including criteria and metrics considered for the selection of the suite.
- In Section 5.3 the benchmark procedure is described.
- Section 5.4 describes the cost performance analysis.

5.2 Benchmark framework

5.2.1 Benchmark categories

The following main categories were considered in selecting the applications composing the benchmark suite. They represent various workloads of particular interest for the European and Italian user community that will mostly benefit from the targeted procured system:

- **Synthetic kernels**. These applications allow to evaluate system-wide performance, independently of specific workloads.
- **Training tasks**. Benchmarks that represent a challenge for the targeted system configuration and that cover a diversity of user workloads.

5.2.2 Benchmark suite

The benchmark suite is composed by the following training tasks, which are a subset of the available MLPerf training benchmark suite, and by the High Performance Linpack benchmark (HPL).

#	Benchmark	Area	type	Short description
1	**GPT-3**	Language: Large language model	Training task	GPT-3 is an autoregressive language prediction model that uses deep learning to produce human-like text
2	**RetinaNet**	Vision: object detection	Training task	RetinaNet is a single-stage object detection model that uses a focal loss function to deal with class imbalance during training.
3	**Stable Diffusion v2**	Marketing, Art, Gaming: image generation	Training task	The Stable Diffusion 2.0 is a robust text-to-image model trained using a brand-new text encoder that greatly improves the quality of the generated images.
4	**HPL**	Computing performance	Synthetic kernel	HPL benchmark solves a linear system of equations of order n, measuring the sustained performance of the whole system.

5	**HPL-MxP**	Computing performance	Synthetic kernel	HPL-MxP provides a combination of LU factorization at a lower precision accuracy and iterative refinement performed afterwards to bring the solution back to 64-bit accuracy. The benchmark seeks to highlight the emerging convergence of high-performance computing (HPC) and artificial intelligence (AI) workloads

Table 5: Training tasks and synthetic kernel composing the benchmark suite.

5.2.3 Metrics

The following metrics are considered for the training tasks of the benchmark suite.

- **Time-to-train (latency in minutes):** the wall-time spent to complete the training task against a defined quality target. Each training task benchmark provides a quality metric and a threshold to determine when the training task is complete.

For the synthetic kernels, the following metrics are considered:

- **Rmax (PFlop/s):** Maximum LINPACK performance achieved for the proposed solution configuration.
- **HPL-MxP (PFlop/s):** Maximum HPL mixed precision performance achieved for the proposed solution configuration.

5.3 Benchmark procedure

In this Section we provide the set of rules to be followed during the execution of the benchmarks, the information on how to execute the benchmarks and the reference values.

5.3.1 Benchmark rules and execution

Regarding the training tasks benchmarks, the rules, models, and input data sets provided in:

https://github.com/mlcommons/, and

https://github.com/mlcommons/training_policies/blob/master/training_rules.adoc

apply. In particular, the benchmarks must be carried out following the **Closed Division approach** described in the training rules (at the link above) that requires using the same preprocessing, model, training method, and quality target as the reference implementation.

Regarding the synthetic kernel HPL benchmark, the rules for qualification to TOP500/GREEN500 lists apply:

https://top500.org/project/call-for-participation/.

Finally, for the synthetic kernel HPL-MxP the following rules apply:

https://hpl-mxp.org/rules.md

5.4 Cost performance analysis

5.4.1 Cost analysis

This section describes the rationale behind the Cost Performance Analysis methodology applied for the LISA system procurement. The "Cost performance Analysis" (CPA) involves projections of OPEX and system performance with the aim to provide a framework for a fair comparison of the offers. For this CPA model the Procurer provides an approximated expression of the total cost of ownership (TCO)

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

Where:

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

In the formula above the operating expenses (OPEX) are approximated - for the purpose of evaluating the offers - to the sole energy cost contribution (en), which for this procurement is estimated to account for most of the operating expenses. Even if the formula above can be considered a rough approximation since multiple costs concur to the total OPEX, the simplicity of the formula carries some strong advantages:

- It is very simple and understandable.
- It lowers the number of factors and variables in play, so simplifying the evaluation process (both for the Procurer and the Candidates).
- Involve the energy cost which is typically the major contribution of the operating expenses and more directly connected to the actual system configuration. Lowering this expense means lowering the energy footprint of the system.
- It reflects directly the architecture and technology of the offered system solutions, making it ideal for evaluating the potentiality of the offered technologies.

Moreover, considering the $CAPEX_{(aq)}$ invariant[6], the formula can be simplified as:

$$TCO = f(OPEX_{(en)})$$

Meaning that the TCO will mainly depend on the $OPEX_{(en)}$ contribution. Since it is not easy to identify *a priori* the AI workloads that will contribute the most to the system load, due to fast evolution of AI applications, the Procurer opts to estimate the $OPEX_{(en)}$) by assuming an average load factor. Moreover, considering a fraction *x* of the offered solution dissipated via DLC[7], therefore adopting the following formula:

$$\mathrm{OPEX(en)} = Peak.Power.Consumption * Load.Factor * Life.Time * Price.Energy * (x * PUE_{DLC} + (1 - x) * PUE_{AC})$$

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

$Load.factor$	Factor applied to scale the power consumption to normal operating conditions rather than HPL	0.6
PUE_{DLC}	Estimated yearly Average Power Usage Effectiveness	1.1

[6] This holds true for fixed price procurement as in this case for the LISA procurement.

[7] Typical ratio can range from 70%-30% (x=0.7) to 90%-10% (x=0.9)

PUE_{AC}	Estimated yearly Average Power Usage Effectiveness	1.3
Life.Time	Lifetime of the system in hours	35040
Price.Energy	Energy cost in Euro/kWh	0.20[8]

Values of Peak.Power.Consumption will be provided according to Section 5 of the Lisa Response Template document and used in the above formula to estimate $OPEX_{(en)}$.

To obtain a simple metric the following formula will be applied for each offer:

$$V_p = \frac{M - O}{M - m}$$

Where:

- O: is the $OPEX_{(en)}$ obtained using the formula above and the committed a Peak.Power.Consumption.
- m: is the minimum value of $OPEX_{(en)}$ corresponding to a power consumption in operations (Load.factor*Peak.Power.Consumption) of 600 kW IT and *x*=1.0.
- M: is the maximum value of $OPEX_{(en)}$ corresponding to a power consumption in operations (Load.factor*Peak.Power.Consumption) of 1200 kW IT (see Section 3.3.2) and *x*=0.0.
- V_p: is the value obtained applying the formula and will be comprised between 0,1.

In case the committed Peak.Power.Consumption would lead to a power consumption in operations (Load.factor*Peak.Power.Consumption) lower than 600 kW IT, V_p will be set to 1.00. In case the committed Peak.Power.Consumption would lead to a power consumption in operations (Load.factor*Peak.Power.Consumption) higher than 1200[9] kW IT, V_p will be set to 0.00. In case Peak.Power.Consumption is not provided the cost-performance score P_C of section 5.5.3 will be set to 0.00.

5.4.2 Performance analysis

The Candidate will provide committed High Performance Linpack (HPL) and High Performance Linpack mixed precision (HPL-MxP) values for the offered solution according to Table 6.

For the HPL and HPL-MxP, the following formula will be applied:

$$V_i = \frac{O_i - m_i}{M_i - m_i}$$

Where:

- i: indicates either the HPL or the HPL-MxP performance metric
- O_i: is the offered committed value for the *i-th* performance metric.

[8] Price per kWh in 2024.

[9] Please note that this is not allowed. According to Req. 3.3.2-1 the offered system must draw in operation conditions less than 1200 kW IT.

- m_i: is the minimum performance according to Table 6.
- M_i: is the maximum performance according to Table 6.
- V_i: is the value obtained applying the formula and will be comprised between 0 and 1 for each performance metric (HPL and HPL-MxP).

To summarize:

i	Min (m_i)	Max (M_i)	Unit of measure
HPL	45000	65000	TFlops
HPL-MxP	300000	650000	TFlops

Table 6: Minimum and maximum values of HPL and HPL-MxP

If the offered committed value (O_i) is not provided for a performance i, the V_i will be assumed to be 0.

5.4.3 Evaluation formula

For each Candidate C, a cost-performance score P_C will be evaluated according to the following formulas:

$$\rho = V_p + V_{HPL} + V_{HPL-MxP}$$

$$P_C = \frac{\rho}{\mathrm{MAX}(\rho_C)}$$

The Candidate can design the offered solution to provide an optimal HPL/HPL-MxP power consumption compromise, to maximize the score associated with this evaluation element. This configuration must be easily replicable during the production phase of the system with the tools and technologies provided in the offer and available to CINECA system administrators.

6. Maintenance and infrastructure availability

The Offer for the procured infrastructure must include a maintenance and support service that ensures high availability as described in Section 6.3, and stability of the procured infrastructure. The term Supplier refers the Candidate awarded for this procurement.

6.1 Maintenance and support requirements

Req.	Description	Category
6.1-1	*Maintenance and support duration* The Candidate will offer maintenance and support for the offered solution for four years.	MRQ
6.1-2	*Compliancy with Leonardo maintenance and support service* In any case, Lisa installation, maintenance and support operations must not compromise the maintenance and support services covering Leonardo. The Supplier of Lisa will be the sole responsible and the sole point of contact for the maintenance and support service of the entire offered solution according to req 6.1.3. This may require an agreement with the provider of Leonardo maintenance and support service, in order to cover requests of support and maintenance on parts of Lisa interfacing the two systems (e.g., datalake front end nodes, InfiniBand L1 switches, and so on.).	MRQ
6.1-3	*Maintenance and support coverage* The maintenance and support will cover all the key hardware and software components of the procured infrastructure, including firmware and all offered programming environment software. This includes all infrastructure components (e.g., racks and power supplies) and network components except for those components provided by CINECA. The Candidate will describe in the Offer all the components that are 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., 5×8, will be covered. The Supplier must ensure the provision of the maintenance and support services, even if there is a dispute with CINECA.	MRQ
6.1-4	*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 by the workload manager are used for system operation, the workload manager must be fully supported by the Supplier.	TRQ
6.1-5	*Reaction times* The Supplier guarantees an appropriate reaction time upon hardware and software issues.	MRQ
6.1-6	*On-site stock*	MRQ

	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 (2-3 hours). The Candidate will provide a list of the intended spare parts included in the on-site stock.	
6.1-7	*On-site support* The Supplier will include one full time equivalent (FTE) position, based on one or multiple qualified persons, to ensure permanent on-site system support during working hours. The on-site 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 on-site persons, a reasonable team size must not be exceeded, and an appropriate coverage of the relevant support fields must always be ensured.	MRQ
6.1-8	*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
6.1-9	*Data deletion* The Supplier will ensure that all client data stored on any, user accessible, non-volatile storage component (incl. 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 methods to ensure the confidentiality of the data stored on non-volatile storage components (e.g., destruction by the customer) may be proposed.	TRQ
6.1-10	*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
6.1-11	*Escalation management process* The Supplier will provide an escalation management process to manage problems priority and critical/non-critical issues.	MRQ
6.1-12	*Regular maintenance and support meetings* CINECA intends to host regular (up to eight meetings per year) face-to-face meetings, to discuss the state of the installation and address any problems. The Supplier will ensure the availability of the necessary (travel) funds required for the attendance of the key support personnel in these meetings.	MRQ
6.1-13	*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 7.	
6.1-14	*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
6.1-15	*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. 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.	MRQ
6.1-16	*Maintenance and support service regulation* This service is considered "*a corpo*"[10] and applies to all the products that are acquired by the Procurer as part of this procurement. The maintenance and support service must include all activities required to ensure regulatory adjustments to software and equipment with reference to all European, national, and regional regulations. All goods included in the service at its launch, even repaired or replacing parts, must comply with current regulations and their evolution. All maintenance and support service interventions must be properly documented. The Supplier or their agent is required to provide the necessary technical assistance, strictly respecting the conditions and the intervention times defined in the specifications. The Supplier responds of the professionalism of the technicians in charge. All parts provided must bear the CE mark and comply with current technical and safety regulations or any regulations issued subsequently, in particular those issued by the UNI and the CEI (Italian Electro technical Committee). The Supplier must specify the compliance of its systems with the applicable safety and emission regulations and electromagnetic compatibility at the time of their offer. In particular, the Supplier must issue a Declaration of Conformity to Law no. 46 - "Safety Standards for Installations".	MRQ
6.1-17	*Maintenance periodic reporting* Periodic maintenance and maintenance activities must be reported, including: • Ticket lists issued by the call centre, including the relevant details. • 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.	MRQ
6.1-18	*CINECA relation with the Supplier*	MRQ

[10] The typically means the service included is to be considered as a complete package.

	Upon awarding the contract and for the conclusion of the contract, the Supplier must nominate a representative to manage all relations with CINECA. The Supplier's representative is the point of contact for any issues that CINECA considers unresolved within the normal relationship with the Supplier (sales manager, technician, call centre, etc.). The Supplier's representative will participate, if required, in regular meetings along with its representatives to update the status of the contract and to share any corrective action needed to comply with the contract. The representative will also be responsible for providing CINECA with the whole documentation necessary for correct access and the use of maintenance and support service (access credentials, etc.). The Supplier's representative must have appropriate professional qualifications and must be available before the supply contract is signed.	
6.1-19	*Documentation requirements* The Candidate will describe the support workflow for software and hardware failures, including information about replacement part logistics and SLAs.	MRQ

6.2 Licenses

Req.	Description	Category
6.2-1	*Licenses* Where applicable, the Candidate must provide licenses for all offered software for the complete duration of the maintenance and support time frame. The software packages provided by CINECA are excluded from this.	MRQ
6.2-2	*Licenses' list* Candidate must provide the complete list of all applicable licenses provided with their quantities	MRQ

6.3 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)*. They are calculated with the following formulas:

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

Where:

- a_i is the availability of each single node "i" (front-end and compute nodes).
- 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.

Req.	Description	Category
6.3-1	*Targeted monthly availability* The design of the procured infrastructure architecture and the maintenance service must aim for a monthly availability of the supercomputer of 95% for the first 3 months of operation and 97% for the remaining of the operational period. For the availability per system partition (percentage of nodes available per partition), the infrastructure architecture and maintenance service must be designed to achieve a monthly availability of 75% for the first 3 months of operation and 85% for the remaining of the operational period.	TRQ
6.3-2	*Minimum monthly availability* The design of the procured infrastructure architecture and the maintenance service must aim for a minimum monthly availability of 85%	MRQ

7. Installation and acceptance

7.1 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. With the term Supplier it is referred to the Candidate awarded for this procurement.

7.1.1 System Installation

Req.	Description	Category
7.1.1-1	*Project Management* The Supplier will provide project management resources for the system installation.	MRQ
7.1.1-2	*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. The Procurer plans to include the system in the TOP500, GREEN500 and HPL-MxP rankings.	MRQ
7.1.1-3	*Installation Time* The procured infrastructure will be installed and accepted according to the timeline defined in Section 3.2.2. The Candidate will describe in the installation plan according to Req. 10-1 and provide the necessary guarantees to meets these targets.	TRQ
7.1.1-4	*Best practices for 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.	MRQ

	• 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. • 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.	

7.1.2 Supply and installation project

Req.	Description	Category
7.1.2-1	Installation project plan The Supplier is responsible for creating a supply and installation project according to Req .10-1 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 Leonardo or CINECA infrastructure. The project must include: • Details of the offered configuration and integration in the CINECA computing system architecture, including setup and interconnection schemes. • Details of hardware and software installation plan, configuration and optimization of the components and partitions. • Details on the interaction with CINECA personnel.	MRQ

	• Implementation plan for procured infrastructure acceptance (see Section 7.2). All interactions with CINECA staff, all training activities, as well as the documentation produced within the project, can be in Italian or English.	
7.1.2-2	*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
7.1.2-3	*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
7.1.2-4	*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

7.2 Acceptance procedure

7.2.1 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 the acceptance tests directed to verify a committed functionality and a performance value included in the Offer are non-negotiable.

7.2.2 Execution of acceptance tests

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

For the acceptance procedure and the verification of the committed benchmark results, the following rules apply:

Req.	Description	Category
7.2.2-1	*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	MRQ

	(WLM) 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 and MPI implementation. If the Offer includes multiple compiler suites or MPI implementations, the Supplier may choose a different combination for each benchmark. All tests will be performed using the production WLM.	

7.2.3 Provisional acceptance tests

7.2.3.1 Hardware checklist

Req.	Description	Category
7.2.3-1	*Hardware checklist* Completeness and consistency of the delivered and installed hardware will be checked against the Offer.	MRQ
7.2.3-2	*Failure thresholds* The thresholds for defective components described in the following table below 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 dysfunctional. • Frontend nodes: all nodes must be functional. • Service nodes: all nodes must be functional. • Ethernet links: less than 0.1% of the links may be dysfunctional. • High speed interconnect links: less than 0.1% of the links may be dysfunctional.	MRQ

7.2.3.2 Software Checklist

Req.	Description	Category
7.2.3.2-1	*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

7.2.3.3 Functional Tests

Req.	Description	Category
7.2.3.3-1	*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	MRQ

	performance level as described in the 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 if present.
- 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 will be verified (component stress tests, see Section 7.2.3.4).
- Verification of single functional component performance level (node, rack).
- Verification of system level performance (Benchmark suite, IO partition, see Sections 7.2.3.5-7.2.3.8),
- Verification of software specifications and offered features.
- Verification of other commitments made in the Proposal.

7.2.3.4 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 all nodes in parallel for 30 minutes without failure.
- A memory stress test will be performed for 24 hours 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.

7.2.3.5 Application and Synthetic benchmarks

The synthetic and application benchmarks included in the benchmark suite (see Section 5.2.3) will be executed with the baseline values as provided by the Supplier in the Offer.

Req.	Description	Category
7.2.3.5-1	*Validation of the committed benchmark results* For the benchmark tests to be considered passed, all committed benchmark results must be achieved within a 5% of relative tolerance.	MRQ

7.2.3.6 Rules for training task benchmarks

The Supplier will dedicate a folder in the system to collect all the input files (application, libraries execution commands, and output files) required to verify and in case replicate the results.

7.2.3.7 Rules for HPL and HPL-MxP benchmarks

The HPL and HPL-MxP benchmarks will be executed on the compute partition according to the TOP500 list and HPL-MxP rules, respectively. During the LINPACK benchmark, the power consumption will be measured according to the GREEN500 run rules. The performance of the benchmarks must be equal or higher (or lower in case of the power consumption) than the performance figure committed to by the Supplier.

7.2.3.8 I/O Performance

Req.	Description	Category
7.2.3.8-1	*Validation of the committed I/O results* The I/O performance of the scratchpad file system will be measured using IOR. The committed I/O performance must be achieved within the relative tolerance of 5%.	MRQ

7.2.4 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).

Req.	Description	Category
7.2.4-1	*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 - as descripted in Section 6.3 - that must result in 85% or above.	MRQ

7.2.5 Final acceptance

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

8. EU added value

Req.	Description
8-1	*Reinforce digital technology supply chain* The tenderer must provide a description on how the offer for the system will reinforce digital technology supply chain in the EU, in hardware and software. For example, by using an EU-based facility for the production of system components, by contributing to the development of EU-based Intellectual Property, etc.
8-2	*Level of integration* The tenderer must provide a description on which level of integration of European technologies, including uptake of existing or upcoming R&D results stemming from EU-funded R&D programs or funded from R&D programs of the EuroHPC participating States.

9. Financing of the contract

The co-funding bodies for this procurement are the European High Performance Computing Joint Undertaking ("EuroHPC JU") for the Union contribution and the Italian Ministry of University and Research ("Ministero dell'Università e della Ricerca – MUR") for the contribution of the Participating State (Italy) where the Hosting Site is established. In accordance with Art. 15 of the Regulation (EU) 2021/1173 on establishing the European High Performance Computing Joint Undertaking[11] the Union contribution shall cover up to 35% of the acquisition costs. While the Union contribution for this contract will be covered by the Digital Europe Programme (DEP), the Participating State contribution will be covered by the budget of the Italian "Piano Nazionale di Ripresa e Resilienza - Recovery and Resilience Fund" PNRR/RRF funds of the Hosting Entity with relevant appropriations.

The contract shall be co-financed by appropriations from DEP amounting up to 9.858.780 EUR excluding VAT and appropriations from the PNRR/RRF – Project No. CN_00000013 – CUP D56G22000380006 "National Centre for HPC, Big Data and Quantum Computing – HPC" under Mission 4, Component 2, Investment 1.4 of the Italian PNRR/RRF funds amounting to 18.309.162 EUR excluding VAT. The final budget, and thus contributions from the Union and Participating State, will depend on the budget proposed in the offer selected for funding. All invoices will be issued to the EuroHPC Joint Undertaking.

To comply with the national RRF/PNRR guidelines all associated invoices must be issued and the respective payments executed before 29th August 2025.

In order to comply with the requirements stemming from the Italian "Piano Nazionale di Ripresa e Resilienza - Recovery and Resilience Fund" PNRR/RRF funds, the applicants are requested to submit the documents listed below together with the applications. This requirement is mandatory.

Req.	Description	Category
9-1	*RRF conditions* The tenderer confirms to fully acknowledge the content of the Annex 10 "Specific RRF conditions"[12].	DOC

[11] OJ L 256, 19.7.2021, p. 3–51.

[12] Annex 10 added on the request of the Hosting Entity to comply with national RRF/PNRR requirements.

10. Documentation

All documents relevant to the bid must be provided in machine-readable electronic form.

Req.	Description	Category
10-1	*Installation plan documentation* The tenderer must provide comprehensive documentation covering all stages of installation and maintenance of the respective systems.	DOC
10-2	*Milestones' documentation* With the completion of each installation project milestone, the following documentation must be provided in PDF and MS Office format: • technical sheets and manuals of all elements installed, • any existing Manual updates adapting them to the new component installed or updated.	DOC

10-3	*Documentation after installation*
	
	The tenderer commits to submitting a digital documentation at the end of the installation in PDF and MS Office format, providing the following information:
	• general description of the solution components,
	• diagrams of the final connections made,
	• explanation of hardware procedures for:
	  ○ Start-up,
	  ○ Shutdown,
	diagnosis in the event of hardware problems.
		DOC

--End of Document—