

CAPITOLATO SPECIALE DESCRITTIVO E PRESTAZIONALE

Premessa

Il LaMMA deve acquisire un sistema di calcolo ad alte prestazioni nell'ambito del progetto Profumo Demonstrator, al fine di ottimizzare le tempistiche di elaborazione dati della catena modellistica meteo marina per fornire servizi di routing operativi per la comunità marittima mediterranea. Ai fini della presentazione di un'offerta valida da parte dei concorrenti, è necessario essere consapevoli delle caratteristiche specifiche del luogo dove andrà installato il sistema richiesto. Dovrà pertanto essere effettuato un sopralluogo nel Data Center del Consorzio LaMMA, dove la fornitura verrà installata, secondo le modalità previste nel disciplinare di gara.

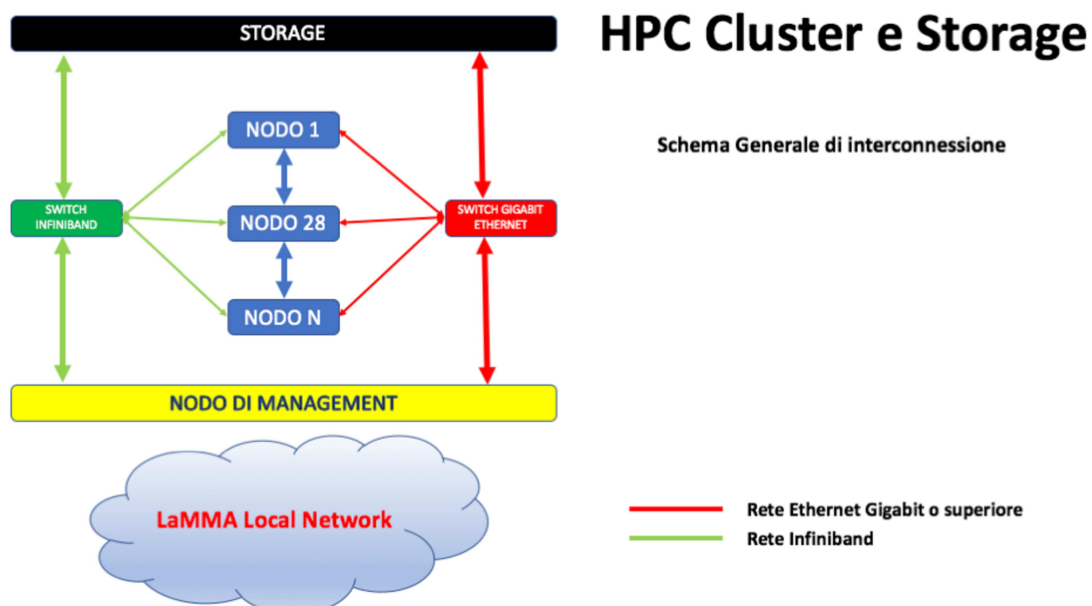
Art. 1 – Elenco delle prestazioni e modalità di esecuzione delle stesse

Il Sistema che il LaMMA intende acquisire consiste in un sistema articolato di componenti, che vanno da sistemi server allo storage, passando per le apparecchiature di rete, atto ad eseguire software di calcolo parallelo su architettura Intel X86_64, utilizzando la comunicazione MPI attraverso le interfacce Infiniband opportunamente configurate.

In particolare, è prevista la fornitura di Nodi ad alte prestazioni (HPC) caratterizzati da architettura dual socket CPU con interfaccia di rete tradizionale ethernet ad alta velocità per la comunicazione di gestione, e connettività ad altissime performance e bassa latenza Infiniband di ultima generazione. Il suddetto Sistema dovrà essere connesso sulle reti dell'Ente tramite il Nodo di Management che dovrà essere a sua volta interconnesso a tutti i dispositivi oggetto della presente fornitura dovendo utilizzare, tra gli altri anche protocolli di comunicazione come l'IPMI (out-of-band management), al fine di controllare le funzionalità e l'operatività dei sistemi.

La connettività Infiniband tra tutti i nodi, compreso quello di management e lo storage, dovrà essere dedicata al traffico dati dei modelli di calcolo.

Una rappresentazione schematica del sistema nel suo complesso e delle connessioni di rete è la seguente:



Obiettivi di performance

L'obiettivo primario della fornitura, è quello di ottenere il miglior risultato possibile di velocità di esecuzione del modello WRF (WEATHER RESEARCH AND FORECASTING MODEL) attraverso il miglior compromesso tra ottimizzazione software e scelta hardware dei componenti.

WRF è un sistema di previsione numerica di mesoscala di ultima generazione, ed è lo strumento software di calcolo principale che verrà utilizzato ai fini del raggiungimento degli obiettivi di progetto.

Per raggiungere questo obiettivo, due sono i punti critici della fornitura:

- Performance di scalabilità del modello WRF.
- Performance del sistema di storage dei dati.

Per entrambi dovranno essere inseriti nell'offerta i valori risultanti da test specifici seguendo le istruzioni dell'**Allegato A**.

Vengono richiesti, in sostanza, i tempi di esecuzione del modello testato dall'operatore economico, la velocità e capacità di gestione del sistema di storage nel trasferimento di varie tipologie di dati.

Obiettivi di qualità e ottimizzazione gestione

Per quanto riguarda gli obiettivi di gestione del sistema, i principali sono:

- consentire la gestione remota di tutti i componenti (con protocolli tipo IPMI);
- compattezza delle soluzioni adottate per minimizzare gli spazi occupati;
- potenza di calcolo adeguata a offrire prestazioni elevate con il modello matematico WRF V3.9.1.1 (WEATHER RESEARCH AND FORECASTING MODEL) i cui riferimenti possono essere acquisiti al seguente indirizzo : <https://www.mmm.ucar.edu/weather-research-and-forecasting-model>)
- basso consumo elettrico, preferendo soluzioni più efficienti dal punto di vista energetico con conseguente minore dissipazione di calore in relazione alle potenze di calcolo erogate;
- nessuna delle componenti hardware della soluzione proposta può essere stata dichiarata End of Life dal Costruttore. Il materiale dovrà essere nuovo di fabbrica: sono pertanto vietati elementi ricondizionati, dimostrativi, da campionari, da fiere, da test interni o in visione presso clienti e in generale tutto ciò che non sia al primo utilizzo.
- Affidabilità hardware dei sistemi

1.1 Fornitura

La fornitura dovrà comprendere il trasporto e la consegna al piano di tutto il materiale previsto.

Di seguito viene elencato il sommario della fornitura richiesta, che verrà poi specificata nei requisiti minimi:

| Descrizione |
|---|
| 1.1.1 Nodi di calcolo |
| 1.1.2 Nodo di management |
| 1.1.3 Sistema di storage con filesystem parallelo |
| 1.1.4 Layer ETHERNET di management |
| 1.1.5 Layer Infiniband |

1.1.1 Nodi di Calcolo

Il nodo di calcolo dovrà essere un sistema server ad alte prestazioni, con architettura x86_64 Intel da inserire in un armadio Rack, caratterizzato dalla compattezza e dalla alta densità di processori per spazio occupato. Dovrà inoltre avere una alimentazione elettrica ridondata e capace di mantenere la piena operatività del dispositivo senza diminuzione delle prestazioni in caso di fault (ogni alimentatore deve supportare il sistema a pieno carico o comunque supportare un fault).

1.1.2 Nodo di management

Il Nodo di Management dovrà essere un sistema compatto, ad alte prestazioni, orientato alla capacità di management richiesta, alla affidabilità e con sistema di gestione dei dischi locali resistente ai fault, oltre ad una alimentazione elettrica ridondata e capace di mantenere la piena operatività del dispositivo senza diminuzione delle prestazioni in caso di guasto (ogni alimentatore deve supportare il sistema a pieno carico o comunque supportare un fault).

1.1.3 Sistema di storage con filesystem parallelo

Il sistema di storage è uno degli elementi più importanti ed il fulcro della comunicazione dei dati durante l'esecuzione del modello, nonché delle procedure di elaborazione dei file e dovrà funzionare in modalità "appliance-like", ovvero saranno utilizzabili, tramite i protocolli di accesso per essi definiti e configurabili tramite apposita interfaccia grafica di alto livello, nascondendo i dettagli implementativi e di tuning; nel caso in cui sia necessaria una licenza per il funzionamento della soluzione storage, questa deve essere inclusa e perpetua, ovvero non soggetta a scadenza; Questo dispositivo dovrà avere una alimentazione elettrica ridondata e capace di mantenere la piena operatività del dispositivo senza diminuzione delle prestazioni in caso di fault.

1.1.4 Layer Ethernet di Management

Il Layer Ethernet di Management, consiste in tutte le apparecchiature di rete necessarie alla comunicazione veloce ed affidabile tra tutti i componenti del sistema di calcolo, relativo al traffico di gestione dei sistemi (accesso remoto ai sistemi, monitoraggio, configurazione,etc)

1.1.5 Layer Infiniband

Il Layer Infiniband, è l'elemento di comunicazione principale del flusso dati del software modellistico, pertanto rappresenta un elemento critico del sistema di Calcolo.

Viene richiesta la fornitura di tutti gli apparati necessari alla comunicazione di tipo Infiniband di ultima generazione, ossia almeno EDR 100Gbit/s .

Gli apparati dovranno essere con doppia alimentazione elettrica, possibilità di sostituzione a caldo e con la possibilità di gestione remota della configurazione (web gui, telnet,ssh ,etc).

La connettività tra dispositivi dovrà essere diretta, ossia non dovranno esserci ulteriori dispositivi di rete attivi tra un nodo di calcolo e l'altro, ad esclusione di eventuali transceiver quando necessari.

1.2 Servizi

Alla fornitura dovranno corrispondere anche una serie di servizi finalizzati alla messa in operatività nel minor tempo possibile del sistema.

Si richiede anche un supporto adeguato e specialistico per poter risolvere le problematiche di gestione avanzate e guasti software nel minor tempo possibile.

Di seguito viene elencato il sommario dei servizi richiesti, che verranno poi specificati nei requisiti minimi:

| |
|---|
| 1.2.1 Installazione del Sistema |
| 1.2.2 Training sull'uso del Sistema |
| 1.2.3 Aggiornamento software e supporto specialistico |
| 1.2.4 Assistenza e Manutenzione hardware |

1.2.1 Installazione del Sistema

Il cablaggio elettrico e di rete dati dovrà essere comprensivo di tutto il materiale necessario all'installazione, al corretto funzionamento di tutti gli apparati in sicurezza, tenendo conto degli assorbimenti elettrici previsti, degli allacci locali da effettuare e dei collegamenti di rete informatica

necessari al funzionamento interno al RACK ed esterno verso la rete dati presente nei locali di installazione.

L'installazione degli apparati forniti dovrà essere realizzata, presso la sede dell'Ente e dovrà comprendere, tra le altre attività:

- a) disimballaggio del materiale;
- b) installazione a rack degli apparati con opportune slitte e sistemi di fissaggio;
- c) cablaggio della parte elettrica a regola d'arte che consenta l'estrazione degli apparati senza compromissione del cablaggio di altre apparecchiature;
- d) cablaggio delle interconnessioni di rete e di comunicazione con lo storage a regola d'arte che consenta l'estrazione degli apparati senza compromissione del cablaggio di altri apparati;
- e) Configurazione sistemistica dell'intera infrastruttura.

I cavi, le slitte ed i materiali necessari per il collegamento e l'installazione negli armadi rack sono a carico del Fornitore e sono considerati inclusi nella fornitura, anche se non esplicitamente indicati nell'offerta.

1.2.2 - Training sull'uso del Sistema

Una volta effettuata la verifica di conformità di cui ai precedenti punti 1.1 e 1.2.1, è richiesta una attività formativa in relazione a tutti gli elementi della fornitura.

Tale formazione dovrà essere focalizzata sulla gestione dei software installati e dei dispositivi con relative interfacce di controllo.

1.2.3 - Aggiornamento software e supporto specialistico

Le prestazioni di aggiornamento software e supporto specialistico dovranno essere garantite per tutta la durata del contratto.

L'amministrazione attiverà telefonicamente specifica richiesta, in relazione alla prestazione ritenuta necessaria.

Entrambe le prestazioni avranno a riferimento l'intero sistema ed a titolo meramente esemplificativo si riportano le ipotesi di seguito elencate:

Aggiornamenti software:

- sistema operativo (minor release, major release)
- moduli del kernel
- patch per la risoluzione di Bug
- software gestione code per la gestione dei modelli matematici
- Compilatori (Intel,PGI,etc)
- Updates firmware apparati di rete e storage

Supporto specialistico:

- Problematiche di configurazione modello WRF con nuovi compilatori di riferimento (Intel,PGI,etc)
- Ripristino delle performance della configurazione di riferimento
- Help Desk per configurazione software delle code di gestione modelli matematici
- Supporto con debug analysis per crash di sistema o dei processi relativi alla esecuzione del modello matematico di riferimento.
- Analisi delle performance e valutazioni per ottimizzazioni e miglioramenti.
- Consulenza generica per ottimizzazione e compilazioni codici sorgenti di modellistica diversi dallo standard WRF

Il servizio dovrà essere erogato in modalità remota, ed in caso di guasto bloccante grave dovrà essere effettuato l'intervento in loco.

1.2.4 - Assistenza e Manutenzione hardware

Il servizio consiste nel mantenimento del perfetto stato di funzionamento di tutti gli apparati per un periodo pari a 5 anni e non deve essere limitato alle componenti hardware ma si deve estendere agli strumenti software correlati con il funzionamento dei dispositivi forniti.

Requisiti minimi

1.1 FORNITURA

| Quantità MINIMA | Descrizione |
|-----------------|---|
| 28 | 1.1.1 Nodi di calcolo |
| 1 | 1.1.2 Nodo di management |
| 1 | 1.1.3 Sistema di storage con filesystem parallelo |
| 1 | 1.1.4 Layer ETHERNET di management |
| 1 | 1.1.5 Layer Infiniband |

1.1.1 Nodi di Calcolo

| Sezione | Elemento | Requisiti Minimi |
|---------------------|-----------------------------|---|
| POWER | Alimentazione | Sistema di alimentazione fornito tale da supportare il consumo a pieno carico (CPU con il massimo Thermal Design Power e massimo numero di DIMM). |
| COOLING | Raffreddamento/Ventilazione | La configurazione del sistema di raffreddamento/ventilazione deve essere tale da supportare il funzionamento a pieno carico con flusso dell'aria fredda aspirata dalla parte frontale (FRONT) ed aria calda espulsa dalla parte posteriore (REAR). |
| MECCANICA | Spazio occupato (in RU) | Min 2 nodi per Rack Unit |
| | Slitte | Slitte per montaggio in armadi 19" compatibili con l'armadio dell'Amministrazione già presente nei locali |
| BOARD (nodo) | CPU | <p>2 Processori di architettura x86_64 con almeno:</p> <ul style="list-style-type: none"> • 16 core fisici per CPU • frequenza di clock di almeno 2.1 GHz. • TDP: 125W <p>Tecnologia: ultima disponibile sul mercato alla data di presentazione dell'offerta</p> <p>Unità di processing logiche quali quelle fornite da Hyper-Threading oppure Simultaneous Multi-Threading non sono conteggiate in questo contesto come core separati.</p> |
| | RAM | <p>192 GiB</p> <p>Memoria di tipo DDR4 con banchi omogenei di almeno 16 GiB di tipo Registered ECC, data transfer con clock minimo 2666 MHz.</p> <p>I moduli di memoria devono essere certificati dal costruttore della scheda madre specificamente per l'utilizzo sulla scheda madre fornita. Tutti i canali di memoria devono essere popolati in maniera bilanciata e certificata dal costruttore della piattaforma, al fine di garantire la massima velocità di accesso alla memoria da parte delle CPU.</p> |
| | SLOT PCIe x16 Gen 3.0 | 1 |
| | Boot | La scheda madre deve supportare il bootstrap via rete con protocollo PXE 2.0 o superiore. Il firmware deve consentire la possibilità di eseguire boot via PXE prima del boot locale. |
| DISCO | DISCO FISSO | 1 disco con interfaccia SATA 6Gbps, capacità 2TB |
| NETWORK | IB | 1 porta 100Gbps EDR con connettore ottico o rame |
| | ETHERNET 1GbE | 2 porte 1Gbps tipologia BaseT onboard module |
| | ETHERNET 100Mb/s | 1 porta 10/100 Mbps BaseT a uso esclusivo per il management del nodo |
| MANAGEMENT | | <p>Deve essere presente un management controller (BMC) compatibile IPMI versione 2.0 o superiore</p> <p>. Il BMC deve consentire almeno il monitoraggio delle ventole (se presenti), della temperatura di CPU e scheda madre, la gestione remota dell'alimentazione elettrica (possibilità di powercycle) e l'accesso criptato alla console seriale attraverso la rete (per esempio via RCMP+ oppure SSH). Il BMC deve consentire la configurazione dell'utente BMC e dei parametri di rete, sia attraverso l'interfaccia web, che tramite un'applicazione a linea di comando in esecuzione locale che funzioni con il s.o. Linux. Il BMC deve mantenere le impostazioni, incluse le configurazioni di accesso e di rete anche qualora l'alimentazione all'unità di sistema venga interrotta; il BMC deve inoltre essere accessibile senza riconfigurazioni una volta che l'alimentazione venga ripristinata.</p> <p>La funzionalità dell'unità di sistema deve essere accessibile sia tramite canale web sia attraverso un'applicazione a linea di comando in esecuzione sui nodi di management. La scheda madre deve supportare la redirectione della console ad una porta seriale accessibile attraverso IPMI Serial-Over-LAN.</p> |

1.1.2 Nodo di MANAGEMENT

| Sezione | Elemento | Requisiti Minimi |
|------------------|-----------------------------|--|
| POWER | Alimentazione | Sistema di alimentazione fornito tale da supportare il consumo a pieno carico (CPU con il massimo Thermal Design Power e massimo numero di DIMM) e la ridondanza su due linee di alimentazione. La rimozione/sostituzione di una delle componenti ridondanti deve poter avvenire a caldo senza alcun impatto sul funzionamento dell'intero sistema. |
| COOLING | Raffreddamento/Ventilazione | La configurazione del sistema di raffreddamento/ventilazione deve essere tale da supportare il funzionamento a pieno carico con flusso dell'aria fredda aspirata dalla parte frontale (FRONT) ed aria calda espulsa dalla parte posteriore (REAR). |
| MECCANICA | Spazio occupato (in RU) | Massimo 2 RU |
| | Slitte | Slitte per montaggio in armadi 19" compatibili con l'armadio dell'Amministrazione già presente nei locali |
| BOARD | CPU | <p>2 Processori di architettura x86_64 con almeno:</p> <ul style="list-style-type: none"> • 8 core fisici per CPU • frequenza di clock di almeno 2.1 GHz. • max TDP: 85W <p>Tecnologia: ultima disponibile sul mercato alla data di presentazione dell'offerta</p> <p>Unità di processing logiche quali quelle fornite da Hyper-Threading oppure Simultaneous Multi-Threading non sono conteggiate in questo contesto come core separati.</p> |
| | RAM | <p>192 GiB</p> <p>Memoria di tipo DDR4 con banchi omogenei di almeno 8 GiB di tipo Registered ECC, data transfer con clock minimo 2666 MHz.</p> <p>I moduli di memoria devono essere certificati dal costruttore della scheda madre specificamente per l'utilizzo sulla scheda madre fornita. I canali di memoria devono essere popolati in maniera bilanciata e certificata dal produttore della piattaforma.</p> <p>La MB deve avere min 12 slot RAM</p> |
| | SLOT PCIe liberi | 3 |
| | Boot | La scheda madre deve supportare il bootstrap via rete con protocollo PXE 2.0 o superiore. Il firmware deve consentire la possibilità di eseguire boot via PXE prima del boot locale. |
| DISCO | DISCO FISSO | Capacità raw di 20TB , interfaccia SATA 6Gbps, capacità dei singoli dischi minima di 4TB 7200 RPM in configurazione RAID hardware 0/1/5/6 con battery backup e cache memory 1GB (non FakeRAID o BiosRAID) |
| NETWORK | IB | 1 porta 100Gbps EDR con connettore ottico o in rame |
| | ETHERNET 1GbE | 2 porte 1 Gbps tipologia BaseT onboard |
| | ETHERNET 100Mb/s | 1 porta 10/100 Mbps BaseT a uso esclusivo per il management del nodo |

| | | |
|-------------------|--|---|
| MANAGEMENT | | <p>Deve essere presente un management controller (BMC) compatibile IPMI versione 2.0 o superiore</p> <p>. Il BMC deve consentire almeno il monitoraggio delle ventole (se presenti), della temperatura di CPU e scheda madre, la gestione remota dell'alimentazione elettrica (possibilità di powercycle) e l'accesso criptato alla console seriale attraverso la rete (per esempio via RCMP+ oppure SSH).</p> <p>Il BMC deve consentire la configurazione dell'utente BMC e dei parametri di rete, sia attraverso l'interfaccia web, che tramite un'applicazione a linea di comando in esecuzione locale che funzioni con il s.o. Linux.</p> <p>Il BMC deve mantenere le impostazioni, incluse le configurazioni di accesso e di rete anche qualora l'alimentazione all'unità di sistema venga interrotta; il BMC deve inoltre essere accessibile senza riconfigurazioni una volta che l'alimentazione venga ripristinata.</p> <p>La funzionalità dell'unità di sistema deve essere accessibile sia tramite canale web sia attraverso un'applicazione a linea di comando in esecuzione sui nodi di management. La scheda madre deve supportare la redirectione della console ad una porta seriale accessibile attraverso IPMI Serial-Over-LAN.</p> |
|-------------------|--|---|

1.1.3 Sistema di Storage con filesystem parallelo

Le tecnologie scelte per lo storage devono essere coerenti con i requisiti posti per il sistema di calcolo. Il sistema di storage deve consentire l'accesso contemporaneo tramite rete infiniband (EDR 100Gbit/s o superiore), 10Gb Ethernet e 1Gb Ethernet.

I nodi di calcolo dovranno accedere allo storage tramite rete InfiniBand.

| Sezione | Elemento | Requisiti Minimi |
|------------------------|------------|---|
| TYPE | Tipologia | Scale out NAS with parallel file system |
| MECCANICA | Dimensioni | Max 6 Rack Unit |
| DISCHI | TIPO | Tecnologia HDD e SSD I dischi SSD dovranno essere utilizzati come cache e/o per la memorizzazione di file di piccole dimensioni, min 64GB di RAM cache volatile |
| | CAPACITA' | Min 60 TB raw |
| NETWORK | InfiniBand | Connessione di tipologia Infiniband verso "Layer INFINIBAND" con velocità massima possibile tale da non far decadere le prestazioni dello storage verso i "Nodi di calcolo" e "Nodo di Management" |
| | ETHERNET | Min 2 porte 10GbE SFP+ , min 2 porte 1GbE |
| FILE SYSTEM | TIPO | Parallel FileSystem : es. GPFS, Lustre, PANFS, BeeGFS con garanzia sulla coerenza della cache dei client |
| RAID | RAID Level | Min Raid 6 (double parity), con tecnologia declustered raid o similare (spare capacity distribuita) |
| Data protection | | Lo storage deve essere dotato di un sistema di protezione in caso di black-out tale da consentire il termine delle operazioni in corso ed effettuare automaticamente un corretto shutdown del sistema |
| Management | | Lo storage deve essere gestibile sia tramite command line che tramite Web GUI |
| Licenza d'uso | Software | Lo storage deve avere licenza d'uso perpetua |

1.1.4 Layer ETHERNET di Management

| Sezione | Elemento | Requisiti Minimi |
|-------------------|-----------------------------|--|
| POWER | Alimentazione | Sistema di alimentazione fornito tale da supportare il consumo a pieno carico (massimo transfer rate) |
| COOLING | Raffreddamento/Ventilazione | La configurazione del sistema di raffreddamento/ventilazione deve essere tale da supportare il funzionamento a pieno carico. |
| | Protocollo di rete | Ethernet 10/100/1000 auto-sensing su tutte le porte di rete |
| MECCANICA | Slitte | Slitte per montaggio in armadi 19" compatibili con l'armadio dell'Amministrazione già presente nei locali |
| NETWORK | ETHERNET 1G | Il totale delle porte usabili deve essere almeno uguale al totale delle porte per il management ETHERNET 1Gb/s dei nodi di calcolo, management e dei sistemi di storage, con tecnologia BaseT e degli apparati di networking. Gli apparati devono supportare la configurazione di VLAN |
| | | |
| MANAGEMENT | | Deve essere presente un sistema di management raggiungibile tramite rete Ethernet compatibile con l'infrastruttura di management dei nodi di calcolo e degli storage. Il sistema deve consentire la completa gestione dell'apparato o dell'insieme di apparati attivi costituenti il layer |
| TOPOLOGIA | | Top Of the Rack |

1.1.5 Layer INFINIBAND

| Sezione | Elemento | Requisiti Minimi |
|-------------------|-----------------------------|--|
| POWER | Alimentazione | Sistema di alimentazione fornito tale da supportare il consumo a pieno carico (massimo transfer rate) e la ridondanza su due linee di alimentazione. La rimozione/sostituzione di una delle componenti ridondanti deve poter avvenire a caldo, senza alcun impatto sul funzionamento dell'intero sistema. |
| COOLING | Raffreddamento/Ventilazione | La configurazione del sistema di raffreddamento/ventilazione deve essere tale da supportare il funzionamento continuo a pieno carico. |
| MECCANICA | Slitte | Slitte per montaggio in armadi 19" compatibili con l'armadio dell'Amministrazione già presente nei locali |
| NETWORK | INFINIBAND | Il totale delle porte usabili deve essere almeno uguale alla somma delle porte IB di tutti gli apparati forniti. Le porte devono di tipo 100Gbps EDR e dotate di connettore ottico o in rame |
| | ETHERNET 1GbE | Almeno uno degli apparati attivi deve essere dotato di 1 porta 10/100/1000 BaseT ad uso esclusivo per il management del layer INFINIBAND. |
| MANAGEMENT | | Deve essere presente un sistema di management raggiungibile tramite rete Ethernet, compatibile con l'infrastruttura di management dei nodi di calcolo e degli storage. Il sistema deve consentire la completa gestione dell'apparato o dell'insieme di apparati attivi costituenti il layer INFINIBAND. Lo switch deve integrare il Subnet Manager |

1.2 SERVIZI

1.2.1 Installazione del Sistema

Il sistema dovrà essere comprensivo di una cluster management suite, dotata di interfaccia a riga di comando, opportunamente configurata, e dei compilatori Intel (www.intel.com) e PGI (www.pgroup.com) con relativo supporto per tutta la durata del contratto.

Il software e l'hardware forniti dovranno essere tra di loro compatibili e i nodi offerti dovranno essere installati e configurati per funzionare per l'uso di uno o più dei seguenti file system paralleli:

- Spectrum Scale (GPFS)
- LUSTRE
- PANFS
- beeGFS

L'operatore dovrà garantire le funzionalità /tool minime di seguito indicate:

1. Bare metal installation completamente automatizzata di sistemi operativi RedHat like (es. CentOS 7.x)
2. Configurazione reti: management (in-band, out-of-band), privata, high performance network per MPI e accesso allo storage
3. Servizi: DHCP, TFTP, NFS, IP forwarding, Gestione utenti centralizzata (LDAP)
4. Gestore di code/risorse: SLURM (<https://slurm.schedmd.com/>)
5. Compilatore Intel® Parallel Studio XE Composer Edition for Fortran and C++ Linux (www.intel.com) con relativo supporto per tutta la durata del contratto
6. Compilatore PGI www.pgroup.com con relativo supporto per tutta la durata del contratto
7. Sistema di Monitoring/Alerting
8. Applicazioni preinstallate (WRF) e ottimizzate, pronte per essere eseguite per le opportune verifiche del collaudo.

1.2.2 Training sull'uso del Sistema

E' richiesto un Training di 8h relativo a tutti gli elementi del sistema.

Dovranno essere illustrate tra gli altri:

- Configurazioni Effettuate sugli strumenti di sviluppo di codice.
- Configurazioni dei servizi principali del sistema operativo
- Gestione delle interfacce CLI e GUI dei dispositivi di rete ethernet ed Infiniband Storage e quant'altro fornito.
- Gestione del software di gestione code con casi di test
- Configurazioni per gestione allarmi e monitoraggio eventi guasti hardware o software

1.2.3 Aggiornamento software e supporto specialistico

E' richiesto un supporto Specialistico della durata di **200 ore** utilizzabili per tutta la durata del contratto, da utilizzare a consumo.

Nel caso di guasto bloccante grave, dovrà essere previsto un intervento in loco di un tecnico.

Per guasto grave si intende una tipologia di blocco dovuto a casi particolari di malfunzionamenti dovuti a problematiche articolate di difficile individuazione attraverso il supporto remoto, non limitatamente a:

- Impossibilità di far funzionare le catene modellistiche operative a causa di bug non precisamente individuabili o situazioni imprevedibili
- Fault o Dump di eseguibili di sistema o relativi a pacchetti installati e mantenuti per il funzionamento delle catene modellistiche (Es: software gestione code)
- Driver delle periferiche installate non più funzionanti o instabili
- Corruzione del filesystem per cause non imputabili alla stazione appaltante.

Livelli di Servizio

Nella seguente tabella sono riportate le definizioni dei livelli di servizio minime richieste per il Servizio di aggiornamento software e supporto specialistico:

DEFINIZIONE DEI LIVELLI DEL SERVIZIO DI AGGIORNAMENTO SOFTWARE E SUPPORTO SPECIALISTICO (SLA)

| Servizio di assistenza e manutenzione | Descrizione | Copertura del servizio |
|---------------------------------------|--|------------------------|
| NBD Next Business Day | Tempo di intervento e ripristino 12h lavorative in continuità con il giorno lavorativo successivo. | 9.00-18.30 Lun-Ven |

Nella seguente Tabella sono definiti gli indicatori di affidabilità dei Servizi richiesto.

| Tipologia di servizio richiesto | Valori di affidabilità minimi richiesti |
|---|---|
| Tempo di intervento per guasto bloccante Software | NBD |
| Tempo di intervento per guasto non bloccante Software | Entro 3 giorni lavorativi |
| Consulenza, richiesta informazioni, tuning | Entro 3 giorni lavorativi |

1.2.4 Assistenza e Manutenzione hardware

Il tempo di intervento e ripristino è da intendersi come il tempo che intercorre dal momento dell'apertura della segnalazione del guasto attraverso telefonata o email da parte del Consorzio, con immediata successiva notifica per email di presa in carico da parte dell'operatore fino al ripristino completo della funzionalità degli apparati.

Livelli di Servizio

Nella seguente tabella sono riportate le definizioni dei livelli di servizio minime richieste per il Servizio di Assistenza e Manutenzione hardware:

DEFINIZIONE DEI LIVELLI DEL SERVIZIO DI SERVIZIO DI ASSISTENZA E MANUTENZIONE HARDWARE (SLA)

| Servizio di assistenza e manutenzione | Descrizione | Copertura del servizio |
|---------------------------------------|--|------------------------|
| NBD Next Business Day | Tempo di intervento e ripristino 12h lavorative in continuità con il giorno lavorativo successivo. | 9.00-18.30 Lun-Ven |

Nella seguente Tabella sono definiti gli indicatori di affidabilità dei Servizi richiesti.

| Tipologia di servizio richiesto | Valori di affidabilità minimi richiesti |
|--|---|
| Tempo di intervento e ripristino guasto bloccante Hardware | NBD |
| Tempo di intervento e ripristino guasto non bloccante Hardware | NBD |
| Tempo di risposta per richiesta di informazioni Hardware | 3 giorni lavorativi |

Art. 2 – Termini, Avvio dell'esecuzione, sospensione e ultimazione dell'esecuzione

1. Il contratto ha una durata quinquennale a decorrere dalla sottoscrizione; le prestazioni devono essere completate entro i termini di seguito riportati:

- a) Consegna della fornitura: di cui al precedente articolo 1 comma 1.1 entro 60 gg dalla sottoscrizione del contratto
- b) Installazione del Sistema: di cui al precedente articolo 1 comma 1.2.1 deve essere avviata entro 5gg dalla consegna della fornitura e deve essere completata entro i successivi 15 gg. lavorativi
- c) Training sull'uso del sistema: di cui al precedente articolo 1 comma 1.2.2 deve essere effettuato entro 3 giorni dalla verifica di conformità delle due precedenti attività e deve essere della durata di almeno 8h.
- d) Aggiornamento software e supporto specialistico, di cui al precedente articolo 1 comma 1.2.3 dovrà essere eseguito per tutta la durata del contratto con le modalità indicate nei requisiti minimi di cui al punto 1.2.3
- e) Servizi di assistenza e manutenzione hardware: di cui al precedente articolo 1 comma 1.2.4 dovrà essere garantito per tutta la durata del contratto a decorrere dalla verifica di conformità delle prime due precedenti attività.

2. Il Responsabile unico del procedimento svolge le funzioni di direttore dell'esecuzione del contratto e al termine di tutte le prestazioni viene redatto un verbale di ultimazione delle prestazioni in contraddittorio e prendono avvio le attività per la verifica di conformità per il rilascio del Certificato di verifica di conformità.

3. Per l'eventuale sospensione dell'esecuzione della prestazione da parte dell'Amministrazione si applica l'art. 107 del D.lgs. 50/2016.

4. L'esecutore che per cause a lui non imputabili non sia in grado di ultimare le prestazioni nel termine fissato può richiederne la proroga ai sensi dell'art. 107, comma 5, del D.Lgs. 50/2016.

Art. 3 – Obbligo di impresa ai sensi dell'Art. 24 comma 1 Legge Regionale 38/2007

1. Ai sensi dell'art. 24, comma 1, L.R. 38/2007 la Società ha l'obbligo di informare immediatamente l'Amministrazione di qualsiasi atto di intimidazione commesso nei suoi confronti nel corso del contratto con la finalità di condizionarne la regolare e corretta esecuzione.

Art. 4 - Importo stimato

1. L'importo complessivo dell'appalto è stimato in 354.000 Euro, oltre Iva nei termini di legge.
2. Per l'espletamento del presente appalto non sono rilevabili rischi interferenti per i quali sia necessario adottare specifiche misure di sicurezza, e che pertanto non risulta necessario prevedere la predisposizione del "Documento Unico di Valutazione dei Rischi da Interferenze" – DUVRI e non sussistono di conseguenza costi della sicurezza di cui all'art. 23, comma 15, del D.Lgs. 50/2016.

Art. 5 – Verifica di conformità

1. Ai sensi dell'art. 102, comma 2, del D.Lgs. 50/2016, le prestazioni contrattuali sono soggette a verifica di conformità, per certificare che l'oggetto del contratto in termini di prestazioni, obiettivi e caratteristiche tecniche, economiche e qualitative sia stato realizzato ed eseguito nel rispetto delle previsioni contrattuali e delle pattuizioni concordate in sede di affidamento.

2. La verifica di conformità di cui al precedente comma 1 è effettuata dal Direttore dell'esecuzione.

3. Il Direttore dell'esecuzione effettua la verifica di conformità in corso di esecuzione al fine di accertare che la relativa prestazione è stata effettuata, in termini di quantità e qualità, nel rispetto delle prescrizioni previste nel presente documento e negli altri documenti ivi richiamati secondo le seguenti specifiche:

- Verifica della Fornitura e installazione del sistema di cui alle lettere a) e b) di cui al precedente art. 2 entro i 3 gg lavorativi successivi alla chiusura delle attività.
- Verifica della attività di Training di cui alla lett. c) del precedente art. 2 entro i 3gg lavorativi successivi alla chiusura delle attività
- Aggiornamento e Supporto Specialistico di cui alla lett. d) del precedente art. 2 il Direttore dell'Esecuzione effettuerà le verifiche in corso di esecuzione trimestralmente
- Assistenza e manutenzione hardware di cui alla lett. e) del precedente art. 2 il Direttore dell'Esecuzione effettuerà le verifiche in corso di esecuzione al termine di ciascuna annualità.

La verifica di conformità definitiva viene effettuata entro 30 giorni dal verbale di ultimazione delle prestazioni, salvo proroga in caso di necessità di svolgimento di ulteriori attività per la verifica, e entro i successivi 10 giorni viene rilasciato il Certificato di verifica di conformità

4. Successivamente all'emissione del Certificato di verifica di conformità si procede al pagamento dell'ultima parte del corrispettivo contrattuale delle prestazioni eseguite e allo svincolo definitivo della garanzia definitiva prestata dall'esecutore a garanzia del mancato o inesatto adempimento delle obbligazioni dedotte in contratto.

L'Amministratore Unico

Dott. Bernardo Gozzini

ALLEGATO A

**Istruzioni propedeutiche per
Eeguire il benchmark delle performance WRF
e del relativo sistema di storage**

PREMESSA

Le istruzioni seguenti vengono fornite per poter procedere al benchmark del modello e delle performance del sistema di storage come richiesto nel disciplinare di gara.

Restano escluse le informazioni di tuning specifiche e di compilazione del modello WRF-ARW che dovranno essere autonomamente trovate dall'operatore economico in quanto dipendenti dalla soluzione ed architettura che verrà proposta.

| Indice | Descrizione |
|--------|---|
| 1 | Procedura di benchmark dei sistemi di storage |
| 2 | Procedura di benchmark nodi |
| | |

1. Procedura di benchmark dei sistemi di storage

Il sistema di storage dovrà essere configurato come singolo volume, accessibile tramite l'operazione di "mount" con uno dei filesystem tra quelli presenti nei requisiti minimi.

I valori di performance richiesti, come indicato nel disciplinare di gara, sono:

- Prestazioni Read Average
- Prestazioni Write Average
- Prestazioni Metadata Average

Prestazioni Read Average e Write Average

Per poter ricavare i valori richiesti, dovrà essere utilizzato il seguente tool di benchmark:

IOzone

Scaricabile dal loro sito web ufficiale:

www.iozone.org

Per ogni opzione/descrizione/approfondimento delle funzionalità di questo software non esplicitamente espresse in questo documento si rimanda al sito web sopracitato oppure alla documentazione di aiuto in dotazione con esso.

Per ottenere i valori da inserire nelle schede Tecniche dell'offerta, per quanto riguarda il Flusso aggregato complessivo per operazioni sia di Write che di Read sequenziali dovrà essere eseguito il seguente comando:

```
iozone -MCce -rZk -sXg -i0 -i1 -tY --n --m clusterfile.txt
```

Dove :

- i0 : test di Write Average
- i1 : test di Read Average
- Y: Numero di nodi usati nel test. Definito dall'operatore economico, deve essere maggiore uguale ad 8 nodi/ 1 processo per nodo
- X: Dimensione del file. Ogni file deve essere più grande del doppio della grandezza della cache/RAM disponibile nel Sistema:
- $X \geq \{2 * \text{Massima quantità di Cache volatile disponibile}\}$
- Z: Grandezza dei record in kB

Il valore Z dovrà essere definito dall'operatore economico tra 128 e 4096:

$128 \leq Z \leq 4096$

- clusterfile.txt: file di configurazione dei nodi su cui dovrà effettuare il test

```
node1 /mnt/VolumeToTest /usr/bin/iozone
node2 /mnt/VolumeToTest /usr/bin/iozone
node3 /mnt/VolumeToTest /usr/bin/iozone
node4 /mnt/VolumeToTest /usr/bin/iozone
...
nodeY /mnt/VolumeToTest /usr/bin/iozone
```

Dal risultato del test si potranno estrapolare i due valori numerici di performance complessiva in scrittura e lettura (WRITE AVERAGE / READ AVERAGE) espressi in KB/s che dovranno essere inseriti nell'offerta, come da riferimento nel disciplinare di gara nei criteri di valutazione dell'offerta tecnica al punto 2.3 "Prestazioni Read Average" e 2.4 " Write Average"

Prestazioni Metadata Average

Il seguente test misura le performance su filesystem paralleli.

Per poter ricavare i valori richiesti in sede di compilazione delle schede Tecniche di offerta, dovrà essere utilizzato il seguente tool di benchmark:

MDTEST 1.9.3

Scaricabile dal sito web:

sourceforge.net/projects/mdtest/

Nel pacchetto della versione richiesta (1.9.3) è presente il file mdtest.c che andrà compilato con i tool mpi necessari per generare il binario che dovrà poi essere eseguito per effettuare i test come di seguito descritto.

Il test dovrà essere eseguito contemporaneamente su 8 nodi di calcolo con un numero di processi per nodo maggiore uguale a 10, tramite MPI/ssh utilizzando i seguenti parametri:

```
mpirun -f host-mdtest.list -np X -ppn Y /app/util/sbin/mdtest -n 3000 -z 2 -b 3 -i 3 -w 4096 -u -d /test/
```

dove

- X =8
- Y >=10
- /test : mountpoint dove è montato il sistema di storage fornito
- host-mdtest.list : comprende i nomi dei nodi configurati per eseguire il test.

Il risultato del benchmark sarà dato dalla differenza valutata in secondi tra la data di fine e la data di inizio del benchmark, come espresso nelle linee di output

```
-- started at <data e ora> --
```

```
-- finished at <data e ora> --
```

Il suddetto valore espresso in secondi dovrà essere inserito nell'offerta, come da riferimento nel disciplinare di gara nei criteri di valutazione dell'offerta tecnica al punto 2.5 "Prestazioni Metadata Average".

2. Procedura di benchmark nodi

Si dovrà procedere a scaricare il software opensource WRF-ARW in versione 3.9.1.1 dal seguente link:

<http://www2.mmm.ucar.edu/wrf/src/WRFV3.9.1.1.TAR.gz>

Nel caso di problemi potete scaricarlo seguendo le istruzioni al seguente link:

http://www2.mmm.ucar.edu/wrf/users/download/get_source.html

dove si dovrà procedere ad una registrazione per poter poi scaricare il software suddetto.

Una volta scaricato si dovrà procedere alla compilazione del suddetto codice, secondo i parametri specifici e la parallelizzazione MPI che l'operatore economico riterrà più opportuna per avere le migliori performance dal sistema fornito.

Si procede allo scaricamento del pacchetto wrf.tar.gz dall'URL fornito in fase di richiesta attraverso la piattaforma START.

Questo è l'elenco dei file presenti una volta scompattato l'archivio:

```
GENPARAM.TBL
LANDUSE.TBL
MPTABLE.TBL
namelist.input
qr_acr_qg.dat
qr_acr_qs.dat
RRTM_DATA
RRTM_DATA_DBL
RRTMG_LW_DATA
RRTMG_LW_DATA_DBL
RRTMG_SW_DATA
RRTMG_SW_DATA_DBL
SOILPARAM.TBL
URBPARAM.TBL
URBPARAM_UZE.TBL
VEGPARAM.TBL
wrfbdy_d01
wrfinput_d01
```

Da notare che per permettere l'esecuzione della simulazione è necessario modificare il file :

```
namelist.input
```

indicando opportunamente i valori di decomposizione in base al numero di processi nella sezione "&domains":

e.g.

per 32 processi $8*4=32$, sarà necessario specificare

```
nproc_x      = 8,  
nproc_y      = 4,
```

Il *time to solution* può quindi essere ricavato con il seguente comando:

```
$ time mpirun ${MPIOPT} ${CMD}
```

ove

MPIOPT, opzioni da passare al comando MPI.

CMD, comando per l'esecuzione della simulazione:

- *Ad esempio:*

```
./wrf.exe
```

Ai files forniti all'interno del file scaricato wrf.tar.gz basterà aggiungere il file **wrf.exe** risultante dalla compilazione effettuata in precedenza per poter effettuare, quindi, i test richiesti:

Run1: Esecuzione su singolo nodo (tutti i core fisici di tutti i socket di un nodo).

In offerta tecnica dovrà essere inserito il valore "Time to Solution" in secondi come da riferimento nel disciplinare di gara nei criteri di valutazione dell'offerta tecnica al punto 1.3 "Prestazioni di calcolo modello WRF con esecuzione sul singolo nodo in modalità shared memory"

Run2: Esecuzione parallela su 256 core, i processi devono essere allocati in maniera da utilizzare tutte le risorse di un nodo prima di utilizzare ulteriori nodi (fill-up).

In offerta tecnica dovrà essere inserito il valore “Time to Solution” in secondi come da riferimento nel disciplinare di gara nei criteri di valutazione dell’offerta tecnica al punto 1.4 “Prestazioni di Calcolo modello WRF con esecuzione su 256 processori (core) in modalità parallela”