



investiamo nel vostro futuro



Dipartimento Interateneo di Fisica
“Michelangelo Merlin”

**FORNITURA DI UNA STRUMENTAZIONE:
“SISTEMA HPC”**

REFERENTE TECNICO DELLA FORNITURA: PROF. ROBERTO BELLOTTI

IMPORTO COMPLESSIVO DEL LOTTO: € 175.000,00 OLTRE IVA

Firma Roberto Bellotti

e-mail: roberto.bellotti@uniba.it

Data: 30 GEN 2014

Allegato tecnico sistema HPC

(Lotto 3)

1 Specifiche Tecniche del Sistema HPC

Fornitura di 20 nodi di Calcolo da rack con acceleratore grafico (GPU modello NVIDIA Tesla K40 o successiva) e uno switch infiniband 40 Gbit/s (Quad Data Rate) con almeno 36 porte QSFP+ InfiniBand.

1.1 Caratteristiche generali della fornitura

- L'intera fornitura è costituita da un numero di sistemi di calcolo tali da soddisfare la fornitura richiesta. Ogni sistema di calcolo sarà composto da una o più *unità di sistema*.
- Il termine *unità di sistema* si riferisce ad un'unità di calcolo basata su una singola scheda madre; il termine *enclosure* si riferisce a un elemento rack-mountable in grado di contenere una o più unità di sistema.

1.2 Caratteristiche delle Unità di Sistema

1.2.1 Scheda Madre

- Sulla scheda madre deve essere presente un management controller (BMC) compatibile IPMI versione 2.0 o superiore. 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 power-cycle) e l'accesso criptato alla console seriale attraverso la rete (per esempio via RCMP+ oppure SSH).
- La funzionalità IPMI dell'unità di sistema deve essere accessibile sia via web sia attraverso un'applicazione a linea di comando in esecuzione su una macchina Linux remota. Il BMC deve consentire la configurazione dell'utente BMC e dei parametri di rete anche attraverso un'applicazione a linea di comando in esecuzione locale che funzioni sotto Linux.
- Il BMC deve mantenere i settaggi, incluse le configurazioni di accesso e di rete, anche qualora l'alimentazione all'unità di sistema sia interrotta; il BMC deve inoltre essere accessibile senza riconfigurazioni una volta che l'alimentazione sia ripristinata.
- La scheda madre deve supportare bootstrap via rete con protocollo PXE 2.0 o superiore. Il BIOS deve consentire la possibilità di eseguire boot via PXE prima del boot locale.
- La scheda madre deve supportare la possibilità di invocare un menu di selezione del device di boot e di selezionare il device di boot all'accensione.
- Il BMC deve supportare la possibilità di cambiare l'ordine dei dispositivi di boot al successivo reboot, tramite un'applicazione a linea di comando sotto Linux o remotamente tramite la sua interfaccia LAN.
- La scheda madre deve supportare la ridirezione della console ad una porta seriale accessibile attraverso IPMI Serial-Over-LAN. Tutti i cambiamenti di settaggi BIOS, l'invocazione del menu di selezione del device di boot e la selezione del device di boot devono essere possibili attraverso questa porta via rete.
- Non è richiesta la ridondanza di BMC o equivalenti.
- E' richiesta la possibilità di connessione a un sistema KVM (Keyboard, Video, Mouse) attraverso connettori standard (tranne PS/2).

- La scheda madre deve fornire almeno 3 porte SAS 1.0 o SAS 2.0, in alternativa 3 porte SATA 2.0 (AHCI) o SATA 3.0 (AHCI).

1.2.2 Processori

- Ciascuna unità di sistema DEVE (pena esclusione) essere dotata di 2 processori multi-core di architettura x86_64 forniti di estensione EM64T oppure AMD64.
- I processori devono fare parte della famiglia Intel Xeon E5-26xx v2 o della famiglia AMD Opteron 63xx, o di famiglie più recenti.
- Ogni processore deve avere almeno 20 core, unità di core virtuali come quelli forniti da tecnologie come hyper-threading o simultaneous multi-threading vengono considerati come core separati.
- Ogni processore deve avere almeno 15 MB di cache L3.
- Il consumo dei processori deve essere tale da non superare i 7 Watt di assorbimento per singolo core di calcolo (i core virtuali sono considerati core a tutti gli effetti).

1.2.3 GPU

Ciascuna unità di sistema deve essere equipaggiata con un acceleratore grafico dotato di GPU del modello NVIDIA Tesla K40 o successiva.

1.2.4 Memoria

- Ciascuna unità di sistema deve essere equipaggiata con almeno 4 GB di RAM per ogni core fisico o virtuale come sopra definito.
- Ciascuna unità di sistema dovrà essere dotata di memorie del tipo DDR-3 registered ECC ed operanti, nel sistema fornito, ad una frequenza effettiva di almeno 1600 MHz.
- I moduli di memoria devono essere approvati dal costruttore della scheda madre specificamente per l'utilizzo sulla scheda madre fornita.
- I canali di memoria devono essere popolati in maniera bilanciata.
- Non è permesso combinare moduli di memoria con differente dimensione, tipo, velocità o fabbricante.

1.2.5 Storage Locale

- Ogni unità di sistema deve avere una capacità di storage minima di almeno 4 TB.
- Almeno uno dei dischi deve essere di dimensione non superiore ai 2TB.
- Tutti i dischi devono avere settori da 512byte, non sono ammessi dischi con settori da 4096 byte.
- Il sistema controller + dischi deve consentire il caricamento del sistema operativo più avanti indicato, via rete (PXE).
- Il controller deve supportare configurazione RAID1 e RAID10
- Tutti i dispositivi storage devono essere hard disk.
- Tutti i dispositivi di storage devono essere forniti di un'interfaccia SAS 1.0 o SAS 2.0 oppure SATA 2.0 o SATA 3.0.
- Tutti i dischi devono essere certificati dal costruttore per operare continuamente 24 ore al giorno, 7 giorni la settimana.
- Tutti i dischi devono supportare Native Command Queuing (NCQ) o Tagged Command Queuing (TCQ)
- Tutti i dischi devono supportare i self test SMART short ed extended.
- Il controller deve verificare lo stato dei dischi attraverso controlli e contatori SMART.

- La combinazione di dischi e di controller deve essere tale da permettere l'aggiornamento del firmware "in place" utilizzando un eseguibile Linux oppure un'immagine che sia PXE bootable.
- Tutti i dischi devono essere coperti da garanzia del rispettivo produttore di almeno 3 anni.

1.2.6 Connettività

- Ciascuna unità di sistema deve essere fornita di:
 - n. 2 schede Gigabit Ethernet rame integrate con supporto PXE ciascuna dotata di connettore RJ45 per i cavi in rame (1000baseT);
 - n. 1 scheda a 2 porte 10 Gbps con connettore SFP+ con supporto PXE;
 - n. 1 transceiver ottico SFP+ SR multimode;
 - n. 1 scheda a Single-Port QSFP 40Gb/s InfiniBand;
- L'accesso alla console remota IPMI deve essere possibile utilizzando una delle due interfacce di rete presenti, ovvero non richiedere necessariamente l'utilizzo di un'interfaccia di rete separata.

1.2.7 Altre richieste

- Tutti i sistemi oggetto del presente lotto devono essere forniti con identiche caratteristiche hardware, versioni di firmware e di BIOS.
- I cavi interni non devono essere punzonati, danneggiati o in tensione meccanica.
- Le funzionalità della unità di sistema definibili via setting NVRAM (per esempio ordine di boot, configurazione RAID, etc.) devono essere mantenute anche in caso di accidentale perdita di alimentazione elettrica.

1.3 Caratteristiche degli enclosure

1.3.1 Caratteristiche fisiche

- Gli enclosure devono essere compatibili per il montaggio nei rack utilizzati nei siti oggetto della fornitura elencati nel paragrafo 2.2.
- Gli enclosure devono avere dimensioni massime di 4U.
- L'occupazione degli enclosure deve essere tale che siano presenti almeno 2 processori per ogni Rack Unit.
- Gli enclosure devono essere approvati dal costruttore per l'uso con il tipo, dimensione e numero di unità di sistema proposte.
- Unità di sistema ed enclosure devono essere progettati in modo tale che in caso di necessità di rimozione o intervento su un'unità di sistema non venga coinvolta nessuna altra unità di sistema.
- Le guide e tutti gli accessori necessari per il montaggio e l'installazione a rack fanno parte della fornitura, così come tutti i cavi interni ed esterni.

1.3.2 Alimentazione elettrica e raffreddamento

- E' richiesta ridondanza di alimentazione elettrica sugli enclosure. La ridondanza deve essere tale che il fallimento di un numero minore o uguale alla metà degli alimentatori presenti su un enclosure non pregiudichi il funzionamento di nessuna delle unità di sistema installate sullo stesso enclosure, anche quando queste sono utilizzate a consumo elettrico massimo.
- Gli alimentatori devono essere certificati 80 Plus Platinum.

- Gli alimentatori devono supportare i requisiti specificati dal costruttore della scheda madre e quelli specificati per tutti i componenti interni all'enclosure.
- Gli alimentatori devono essere dotati di compensazione di fattore di potenza ($\cos(\varphi)$) secondo lo standard IEC 61000. Per un sistema a pieno carico, è richiesto un fattore di potenza di 0.9 o superiore.
- Devono essere forniti cavi di alimentazione con spine e prese standard IEC di lunghezza idonea e cavi di rete standard UTP categoria 6 di lunghezza da determinarsi in collaborazione con l'Università degli Studi di Bari *Aldo Moro* destinataria della fornitura prima del montaggio delle macchine nei rack. Il numero e la lunghezza di tali cavi, così come il tipo esatto di presa elettrica di alimentazione, dipendono dal tipo di soluzioni tecniche presentate e non possono pertanto essere specificate rigidamente prima dell'aggiudicazione della gara.
- Il sistema di raffreddamento deve essere dimensionato in modo che l'enclosure e tutti i componenti in esso installati vengano mantenuti ad una temperatura compatibile con quanto specificato per il funzionamento del sistema, assumendo un'installazione in un rack pienamente popolato e con una temperatura dell'aria in ingresso compresa tra 15 e 35 gradi centigradi. L'aria calda deve poter essere espulsa solamente attraverso la parte posteriore dell'enclosure.
- I ventilatori, fatta eccezione per quelli delle CPU e degli alimentatori, devono essere tra di loro ridondanti.

1.4 Caratteristiche software e documentazione

1.4.1 Sistema Operativo

- Il sistema operativo da installare sarà Scientific Linux 6.4 o superiore, disponibile gratuitamente dal sito <http://www.scientificlinux.org> (il sistema è compatibile a livello binario con RedHat Enterprise Linux 6 Server x86_64).
- I sistemi non devono contenere componenti o device che richiedano driver non inclusi in tale sistema operativo.
- Si richiede l'installazione del sistema operativo Scientific Linux nell'ultima versione disponibile della serie 6.X al momento della fornitura
- Tutti i computer dovranno essere in grado di fare il boot del sistema operativo citato senza essere connessi a tastiera, video, mouse o console seriali.

1.4.2 Settaggi

- Tutti i sistemi devono essere consegnati con settaggi identici. In particolare, le configurazioni di disco, i livelli di revisione del firmware (BIOS, BMC), i settaggi memorizzati in NVRAM e i jumper settings devono essere i medesimi. L'operatore economico dovrà prendere contatto con i siti per la definizione dei settaggi prima della consegna. I settaggi richiesti possono essere diversi a seconda della sede di installazione.

1.4.3 Strumenti flash linux e immagini BIOS/firmware

- Devono essere forniti strumenti dedicati a linea di comando in grado di effettuare l'aggiornamento di BIOS e di BMC. Questi strumenti devono poter essere eseguiti all'interno del sistema operativo Linux.

1.4.4 Documentazione

- Le unità di sistema e gli enclosure della fornitura devono essere consegnati con un set completo di manuali in italiano o in inglese, sotto forma di CD e in formato PDF oppure HTML. L'operatore economico deve accordare ai partner del progetto RECAS il diritto di ridistribuire tali manuali al proprio personale operativo.
- Prima delle operazioni d'installazione, l'operatore economico dovrà consegnare un file elettronico in formato csv (comma separated variables) contenente il numero seriale di ciascun enclosure e i MAC address di ogni scheda di rete.

1.5 Descrizione dello switch di connessione

- Si intende compreso nella fornitura del sistema di calcolo in oggetto uno switch infiniband da almeno 36 porte con le seguenti caratteristiche:
 - 40Gbit/s QDR (Quad Data Rate) bidirezionale non-blocking per porta
 - supporto a IBTA 1.2.1
 - latenza inferiore ai 200 nsec
 - fattore di forma 1U
 - alimentatore e sistema di raffreddamento ridondanti e rimovibili a caldo
- Dovranno essere forniti anche 36 cavi di collegamento 40 Gb/s QSFP+ compatibili con lo switch e con i nodi di calcolo che verranno forniti.

1.6 Descrizione delle caratteristiche dei sistemi di calcolo

- Per ogni sistema di calcolo dovranno obbligatoriamente essere specificate le seguenti informazioni:
 1. Dimensione e numero degli enclosure
 2. Capacità degli enclosure intesa come numero di processori per Rack Unit;
 3. Caratteristiche del processore offerto;
 4. Numero e caratteristiche dei dischi offerti;
 5. Caratteristiche della memoria e GB per core offerti;
 6. Caratteristiche dell'acceleratore grafico per unità di sistema;
 7. Caratteristiche dello switch infiniband offerto;
 8. Caratteristiche delle schede di connettività offerte;
 9. Misure delle prestazioni di ogni unità di sistema in termini di Hep-Spec: si chiede di indicare la potenza di calcolo di ogni unità di calcolo. Tale valore verrà verificato in fase di verifica di conformità.
 10. Potenza assorbita dai nodi di front-end: dovranno essere indicate la potenza AC apparente (in VA), la potenza AC attiva (in W) e il fattore di potenza sotto pieno carico e in modalità idle;
 11. Potenza assorbita di un enclosure pienamente popolato con le unità di sistema proposte: dovranno essere indicate la potenza AC apparente (in VA), la potenza AC attiva (in W) e il fattore di potenza sotto pieno carico e in modalità idle;
 12. La configurazione delle schede di rete suggerita per il supporto delle funzionalità di PXE e IPMI.

2 Consegna e Installazione delle apparecchiature

2.1 Consegna delle apparecchiature

- Il Fornitore dovrà provvedere alla consegna delle apparecchiature entro 30 giorni solari dal momento in cui verrà sottoscritto il contratto e comunque non prima del 1 maggio 2014. L'intero sistema dovrà essere installato nei rack, reso operativo e validato entro il 31 luglio 2014.
- L'indirizzo di consegna è il seguente:
 - **Dipartimento Interateneo di Fisica “Michelangelo Merlin”**
Campus Universitario

Via E. Orabona n. 4
70125 Bari

2.2 Installazione delle apparecchiature

- Tutto l'hardware fornito dovrà essere installato nei rack, reso operativo e validato ad opera del fornitore.
- L'installazione dovrà comprendere per tutte le apparecchiature la cablatura all'alimentazione elettrica e alla rete, inoltre, per i Sistemi di Storage del Lotto 2, dovrà comprendere il cablaggio tra le unità di storage e i nodi di front-end, il montaggio e la configurazione delle schede Fibre Channel in modo da avere 2 canali ridondati per nodo.
- Il fornitore dovrà inoltre provvedere al settaggio di rete delle console remote dei sistemi.
- Le guide e gli accessori (bulloni e viti varie) necessari per il montaggio e l'installazione a rack, così come i cavi di alimentazione, di rete e le fibre, fanno parte della fornitura.
- I rack utilizzati presso l'Università degli Studi di Bari *Aldo Moro* sono del seguente tipo:
 - Rack della ditta APC, modello SX AR3100
- Le PDU, disponibili presso l'Università degli Studi di Bari *Aldo Moro*, sono del tipo IEC C13.
- L'installazione dovrà essere concordata con un anticipo di almeno 10 giorni, in modo da impedire l'accesso ai locali per altre attività. A tale scopo contattare il Direttore dell'esecuzione del contratto, Prof. Roberto Bellotti (080-5443171, <roberto.bellotti@uniba.it>). Il Prof. Roberto Bellotti sarà garante dell'eliminazione dei rischi da interferenza.