Capitolato Speciale d'Appalto

Fornitura dell'infrastruttura di calcolo, di storage e di networking del Supercomputing Center del CMCC

Sommario

1 Introduzione

Il CMCC intende dar luogo ad un procedimento per l'aggiornamento ed il consolidamento del Supercomputing Center (SCC) del CMCC.
Il presente Capitolato Tecnico disciplina gli aspetti tecnici della fornitura di hardware, software e quanto altro necessario, ivi incluso l'installazione, la manutenzione e l'assistenza tecnica specialistica, per l'aggiornamento ed il potenziamento dell'attuale Data Center del CMCC che ha sede a Lecce.

2 Oggetto della fornitura

L'oggetto dell'appalto riguarda la fornitura dei sistemi, apparati, software e servizi che saranno necessari per la realizzazione del nuovo Data Center del CMCC.

In particolare, si richiede la fornitura di:

- **Sistema di calcolo scalare/parallelo** (HPC Cluster, di seguito **HPCC**);
- **Sistema di storage HPC** (HPC Storage System, di seguito **HPCSS**);
- **Sistema di storage Tier2** (Tier2 Storage System, di seguito **T2SS**);
- **Sistema di archiviazione** (Deep Archive Storage System, di seguito **DASS**);
- **Infrastruttura rete dati** (Core Data Center Network, di seguito **CDCN** e Management Data Center Network, di seguito **MDCN**);

Nell'Allegato A denominato "CMCC Supercomputing Center – Schema generale" viene fornito uno schema completo di alto livello del data center, nel quale sono rappresentanti sia i sistemi oggetto della presente fornitura, sia altri sistemi previsti in future espansioni e pertanto non inserite nel presente capitolato tecnico.

Eventuali altre componenti e servizi, anche se non esplicitamente menzionati ma comunque necessari per la gestione, l'integrazione e il corretto funzionamento dei sistemi forniti (ad es. cavi di collegamento, strumenti HW/SW per la configurazione, per la gestione e per il monitoraggio, firmware, ecc.) dovranno anch'essi essere compresi nella fornitura.
La fornitura, inoltre, dovrà appartenere alla più recente generazione di prodotti rilasciati in commercio ed essere costituita esclusivamente da elementi nuovi di fabbrica.

3 Sistema di calcolo scalare/parallelo (HPCC)

L'obiettivo è quello di realizzare un sistema di High Performance Computing (HPC) basato su di un sistema operativo Linux a supporto delle attività di ricerca del CMCC. Il Sistema sarà formato da un insieme di nodi di calcolo e da un appropriato numero di nodi di servizio.
Il sistema che si vuole realizzare **dovrà avere una capacità di calcolo aggregata di picco (theoretical peak performance) pari ad almeno 1,2 PetaFlops** (1200 TeraFlops) considerando unicamente le **operazioni floating-point in doppia precisione**.

3.1 Architettura del sistema

L'architettura del sistema di calcolo scalare/parallelo dovrà essere del tipo multi-nodo con nodi di calcolo e nodi di servizio (I/O nodes, login node, management e monitoring nodes, etc.).

L'architettura del sistema che s'intende realizzare è quella standard di un cluster e sarà quindi costituita dai seguenti componenti:

- nodi di calcolo;
- nodi di servizio (I/O nodes, login node, management e monitoring nodes, etc.);
- infrastruttura di rete per il management del sistema di calcolo;
- infrastruttura di rete per l'interconnessione veloce e a bassa latenza dei nodi;
- sistema di storage ad alte prestazioni.

3.2 Caratteristiche dei nodi di calcolo

I nodi di calcolo dovranno essere basati su piattaforme altamente integrate e idonee all'ottimizzazione degli spazi, della potenza elettrica assorbita e dissipata.
In particolare, per quanto riguarda la dimensione dei nodi di calcolo, essi dovranno avere un'occupazione non superiore a 0,5U per singolo nodo.

Il singolo chassis che ospiterà i nodi dovrà soddisfare i seguenti requisiti:

- Dimensione del Fault Domain (numero di nodi di calcolo contenuti in uno chassis) non superiore a 4;
- Alimentazione ridondata in modalità 1+1. La caduta di un alimentatore non deve determinare alcuna variazione delle prestazioni e/o della potenza di calcolo generata dai nodi contenuti nello chassis.

Le PDU (Power Distribution Unit), che saranno utilizzate per l'alimentazione dei nodi di calcolo, dovranno essere di tipo *managed* dotate di interfaccia di rete Ethernet e capace di permettere l'accensione/spegnimento a distanza e la misura del carico assorbito da ogni singola presa o almeno da sottogruppi di prese.

Inoltre, i nodi di calcolo dovranno avere le caratteristiche minime descritte di seguito.

3.2.1 Processori

Per quanto riguarda i processori, i requisiti minimi che dovranno essere soddisfatti sono i seguenti:

- ciascun nodo dovrà essere dotato di 2 processori multi-core x86 a 64bit;
- ogni processore dovrà avere un numero di core fisici compreso tra 16 e 24;
- frequenza del processore $\geq$ 2,0 GHz;
- ogni processore dovrà avere almeno 22 MB di cache L3

Per il calcolo delle prestazioni di picco del sistema si procederà nel seguente modo:

- verrà presa in considerazione esclusivamente la frequenza nominale dei processori, escludendo meccanismi di *burst*, *overclocking* o similari;
- dovranno essere presi in considerazione unicamente i nodi di calcolo (escludendo quindi tutti i nodi di servizio).

3.2.2 Memoria

- Ciascun nodo dovrà essere equipaggiato con almeno 96 GB di RAM;
- Ciascun nodo dovrà essere dotato di memorie del tipo DDR-4 *registered* ECC *dual rank* ed operanti, nel sistema fornito, ad una frequenza effettiva di almeno 2666 MHz;
- I moduli di memoria offerti dovranno essere approvati e certificati dal costruttore della scheda madre;
- I canali di memoria dovranno essere popolati in maniera bilanciata ed in base alle indicazioni fornite sia dal produttore del processore, sia dal produttore della scheda madre al fine di ottenere le prestazioni ottimali;
- Non sarà permesso combinare moduli di memoria con differente dimensione, tipo, velocità o fabbricante.

3.2.3 Storage locale

Considerato che lo storage locale dei nodi dovrà ospitare unicamente il sistema operativo, i nodi di calcolo dovranno essere dotati di un disco di piccole dimensioni di tipo SSD e con capacità minima di 240GB.

3.2.4 Board Management Controller (BMC)

Tutti i nodi dovranno essere dotati di un board management controller (BMC) compatibile IPMI versione 2.0 o superiore e Redfish. Il BMC deve consentire almeno il monitoraggio delle ventole (se presenti), della temperatura dei processori e scheda madre, la gestione remota dell'alimentazione elettrica, l'accesso criptato alla console seriale attraverso la rete (ad es. via RCMP+ oppure SSH) e la possibilità di interagire col sistema in modalità virtual console (remotizzazione KVM e virtual media). La BMC dovrà essere dotata di interfaccia di rete almeno 100MBps Base-T separata dalla rete di produzione.

3.2.5 Connettività

Ciascun nodo dovrà essere connesso alla rete di Cluster Management Network (CMN) la cui caratteristiche sono descritte nel paragrafo 3.4

Ciascun nodo dovrà essere dotato di almeno n. 1 scheda di rete per interfacciarsi alla rete Infiniband di interconnessione veloce intra cluster di ultima generazione disponibile. Le caratteristiche della rete Infiniband sono descritte nel paragrafo 3.5.

Un numero di nodi non inferiore al 10% del totale dei nodi forniti dovrà essere dotato di una scheda di rete dual-port per interfacciarsi alla Core Datacenter Network (CDCN) con velocità di connessione a 10Gbps. Le caratteristiche della sono Core Datacenter Network (CDCN) sono descritte nel paragrafo 4.1.

3.3 Caratteristiche dei nodi di servizio

Il sistema dovrà essere dotato di un appropriato numero di nodi di servizio (I/O node, login node, management e monitoring node, etc.).

I nodi di servizio dovranno avere la stessa identica configurazione dei nodi di calcolo (cfr. par. 3.2) tranne che per la configurazione delle seguenti componenti:

- dimensione minima del singolo nodo non inferiore ad 1U;
- storage locale: ciascun nodo dovrà essere dotato di nr 2 hard-disk in configurazione RAID 1 (implementato in hardware) con capacità minima pari a 2TB;
- doppia alimentazione di tipo *hot-swap*.

Inoltre, per quanto riguarda il numero minimo di nodi di servizio e la quantità di memoria, le configurazioni minime che dovranno essere soddisfatte sono descritte nella tabella di seguito riportata:

	Login node	I/O node	Management node	Monitoring node
numero minimo di nodi richiesti	4	Se previsti, il numero di I/O node dovrà essere tale da soddisfare i requisiti minimi richiesti per il sistema HPCSS (cfr. par. 3.6)	2	2
memoria	384 GB	384 GB	192 GB	192 GB

3.4 Cluster Management Network (CMN)

La Cluster Management Network (CMN) sarà la rete dedicata al system management del cluster HPCC. Pertanto, al fine di interconnettere tutti i nodi del sistema HPCC alla rete CMN, si richiede la fornitura di tutti i necessari componenti hardware e software per la realizzazione di una rete di tipo Ethernet con velocità di almeno di 1Gbps.

3.5 Infrastruttura di rete per l'interconnessione veloce e a bassa latenza dei nodi

Al fine di interconnettere tutti i nodi del sistema HPCC, si richiede la fornitura di tutti i necessari componenti hardware e software per la realizzazione di una rete veloce e a bassa latenza.
Tale rete sarà basata su protocollo Infiniband EDR a 100Gbps, con un rapporto di *over-subscription* non superiore a 2:1.

Si motiva la scelta di Infiniband in quanto la soluzione proposta dovrà essere compatibile ed interoperabile con l'infrastruttura di interconnessione a bassa latenza già esistente presso il CMCC (FDR-10 IB), in ottica di riutilizzo delle componenti.

Per la realizzazione di tale rete, le interfacce di rete, gli apparati di switching ed i cavi dovranno essere dello stesso produttore e certificati per il mutuo utilizzo.

La soluzione offerta dovrà includere anche un software di management, gestione e *provisioning* del sistema di interconnessione in grado di raccogliere e visualizzare col massimo dettaglio i dati prestazionali e quelli relativi allo stato di salute del sistema e di effettuare *tuning* e preventive *maintenance* delle componenti.

3.6 Sistema di storage HPC (HPCSS)

Il sistema di calcolo/scalare parallelo (HPCC) dovrà essere dotato di un sistema di storage HPC (HPCSS) dedicato. Lo spazio disco utile fornito da tale sistema sarà utilizzato unicamente come "spazio di lavoro" (scratch file system) per memorizzare temporaneamente i risultati delle simulazioni numeriche che saranno eseguite sul sistema HPC. Per tale motivo, il sistema HPCSS dovrà garantire elevate prestazioni in termini di I/O anche in presenza di un elevato numero di richieste concorrenti.

Il sistema di storage HPC dovrà soddisfare le seguenti caratteristiche e funzionalità:

- capacità utile complessiva pari ad almeno 1 PebiBytes (come definito nel par. 13.2);
- performance aggregate di I/O sul file system pari ad almeno 35 GByte/sec.

Inoltre, **a livello hardware**, il sistema di storage HPC dovrà soddisfare le seguenti caratteristiche e funzionalità:

- possibilità di espandere il sistema in termini di scalabilità verticale o orizzontale;
- compatibilità certificata con la soluzione di cluster parallel file system che verrà proposta;
- assenza di singoli point of failure tramite ridondanza di tutte le componenti critiche;
- possibilità di sostituire a caldo (hot swap) tutte le componenti hardware critiche (unità di alimentazione, controller, etc.);
- gestione dei dischi guasti con sistema a doppia sicurezza di tipo "hot spare" e/o "distributed spare" ed in ogni caso con sostituzioni a caldo (sono ammesse implementazioni software delle funzionalità "spare");
- aggiornamento firmware senza interruzione di servizio;
- tutti dispositivi forniti, inclusi gli hard-disk, dovranno essere di tipo "enterprise", ovvero certificati per l'uso H24;

- gli apparati dello storage dovranno avere funzionalità in grado di garantire l'assoluta consistenza dei dati in caso di fault come, ad esempio, il "destaging" dei dati presenti in cache prima dello spegnimento in caso di assenza di alimentazione elettrica;
- gli apparati dello storage dovranno avere funzionalità di auto-diagnosi o essere interconnessi ad un sistema di rilevamento dei fault, al fine di informare tempestivamente gli amministratori del sistema di eventuali guasti;
- gli apparati dello storage dovranno avere funzionalità di monitoring in termini di utilizzo e performance del sistema HPCSS o essere interconnessi ad un sistema di monitoring esterno.

Infine, **a livello software**, il sistema di storage HPC proposto dovrà soddisfare le seguenti caratteristiche e funzionalità:

- lo spazio utile di lavoro dovrà essere implementato tramite una soluzione di cluster parallel file system come, ad esempio, Lustre, GPFS o equivalente;
- alta affidabilità, prestazioni elevate, parallelizzazione degli accessi e bilanciamento del carico sia a livello dei dati che dei metadati;
- compatibilità, a livello di file system, allo standard POSIX ed ai linux kernel;
- supporto delle funzionalità di:
    - Access Control List "ACL";
    - quota disco a livello utente, gruppo, fileset o directory ;
    - ridimensionamento dinamico dei volumi;
    - gestione dei failover e health monitoring;
    - performance monitoring;
    - esportabilità via NFS.

La soluzione offerta deve comprendere altresì tutte le componenti software e licenze, necessarie a garantire la messa in esercizio e il funzionamento dello spazio disco di lavoro del sistema di calcolo parallelo HPCC, nel rispetto delle funzionalità e dei requisiti minimi sopra descritti.

3.7 Software

Per quanto riguarda invece il software, si richiede la fornitura dei seguenti pacchetti/prodotti.

3.7.1 Sistema operativo

Il sistema operativo da utilizzare sui nodi di calcolo e, più in generale sulle entità del cluster HPCC, dovrà essere un Sistema Operativo Linux x86_64 di ultima generazione di tipo Red Hat o CentOS.

3.7.2 Ambienti di sviluppo

Si richiede la fornitura degli strumenti di sviluppo software di seguito descritti.

Suite per lo sviluppo e il debugging di codici seriali e paralleli costituita da:

- compilatori Fortran, C e C++, perl, python nell'ultima versione disponibile per l'architettura proposta, sia in versione enterprise, sia in versione GNU.
- ambiente software per lo sviluppo e l'esecuzione di applicazioni parallele basate su message-passing MPI e OpenMP con supporto del network di interconnessione del cluster;

- environment modules per la gestione e configurazione dei path di software, librerie e tools HPC.

Gli strumenti forniti dovranno consentire la compilazione e l'installazione delle principali librerie per il calcolo scientifico.

3.7.3 Scheduler e Resource Manager

Un'efficiente pianificazione dei carichi di lavoro può garantire che le risorse di calcolo (e storage) siano utilizzate in modo efficace per accelerare i tempi di elaborazione e di produzione dei dati da parte delle simulazioni numeriche. Di conseguenza, le applicazioni di gestione dei carichi di lavoro devono operare in modo efficiente e rispondere dinamicamente, anche negli ambienti più esigenti.

Pertanto, al fine di aumentare la produttività degli utenti utilizzatori del sistema HPCC ed al fine di migliorare l'utilizzo delle risorse disponibili (in termini di RAM, CPU, ecc), si richiede la fornitura di un software (tipo LSF, PBS o equivalente) per l'implementazione di un sistema di code (batch, pipe, ecc.) che permetta la sottomissione di job seriali/paralleli e la gestione dell'allocazione delle risorse. Tale software dovrà, in particolare, poter garantire:

- massima efficienza e produttività nella gestione dei job utente;
- ottimizzazione statica e dinamica dell'allocazione delle risorse;
- definizione e gestione di un sistema di code;
- gestione di job seriali, paralleli ed ibridi openMP / MPI,;
- gestione delle priorità dei job;
- possibilità di sospendere i job;
- possibilità di implementare differenti policy di scheduling;
- possibilità di garantire priorità a classi di job e/o utenti;
- consentire l'implementazione di workflow complessi di job (catene operative);
- staging intelligente dei dati di input;
- analisi statistica dell'utilizzo delle risorse (accounting su base temporale arbitraria per gruppi, utenti e progetti ed altro);

3.8 Cluster management system (CMS)

La realizzazione del cluster HPCC comporta l'installazione e la messa in opera di numerosi dispositivi tra cui server, compute node, dispositivi di rete ed altro. Diventa pertanto indispensabile dotarsi di strumenti che permettano con facilità il provisioning, il monitoraggio e la gestione del cluster.

Accanto a tali esigenze si aggiunge, anche, quella di dover garantire la continua operatività del cluster fornendo possibili soluzioni di recovery, come ad esempio image revision control e recovery, node failover, redeploy di specifici environment.

Si richiede pertanto la fornitura di un sistema di cluster management, che dovrà:

- interfacciarsi sia con la Cluster Management Network (CMN), sia con la Management Datacenter Network (MDCN);

- essere installato su almeno numero 2 server e operare in modalità active/active o active/passive purché in ogni caso sia garantita la disponibilità del servizio;
- gestire principalmente i dispositivi del cluster HPCC (saranno valutate positivamente soluzioni in grado di gestire dallo stesso sistema gli altri dispositivi oggetto della medesima fornitura "funzionalità multi-cluster");
- fornire un'interfaccia web oriented (GUI) compatibile con i principali browser oltre che un'interfaccia command line (CLI);
- gestire le principali distribuzioni linux come ad esempio Red Hat Enterprise Linux, SUSE, CentOS, Scientific Linux, Ubuntu Server Long Term e fornire funzionalità di gestione degli aggiornamenti;
- fornire funzionalità di monitoraggio completo che consentano agli amministratori di sistema di controllare, visualizzare e analizzare con facilità le risorse a livello hardware e software anche attraverso la definizione di metriche personalizzate;
- essere in grado di rilevare situazioni di allarme o pre-allarme e, secondo modalità personalizzabili, informare gli amministratori di sistema oltre che attivare delle azioni predefinite (triggers alerts e triggers actions);
- fornire una serie completa di strumenti per consentire agli amministratori di sistema di sviluppare, sottoporre a debug e distribuire librerie e codice HPC;
- gestire la scalabilità del cluster secondo policy manuali e/o di "automated scaling";
- fornire funzionalità di node failover, image revision control e recovery;
- interfacciarsi con i principali protocolli in uso nei sistemi di Identity and Access Management (NIS, LDAP e Active Directory).

3.9 Requisiti di ingombro del sistema HPCC

Al fine di minimizzare l'ingombro, lo spazio occupato dai rack che ospiteranno i nodi di calcolo del sistema HPCC non dovrà superare la superficie equivalente allo spazio occupato da 6 rack standard 42U.

3.10 Efficienza energetica del sistema HPCC

Al fine di minimizzare i costi di funzionamento operativo del sistema HPCC (tutti i nodi, storage HPCSS e rete di interconnessione INFINIBAND), si richiede che il valore del PUE, riferito al solo sistema HPCC, non sia superiore ad 1,3 (ovvero PUE $\leq$ 1,3) dove il PUE è definito, in questo caso, come:

$$PUE_{HPCC} = \frac{P_{CDZ-HPCC}}{P_{HPCC}}$$

dove:

$P_{CDZ-HPCC} = Potenza\ necessaria\ per\ alimentare\ e\ raffreddare\ il\ solo\ sistema\ HPCC$

$P_{HPCC} = Potenza\ necessaria\ per\ alimentare\ il\ sistema\ HPCC$

Il concorrente dovrà includere in fornitura una soluzione di raffreddamento e/o contenimento del calore dissipato in grado di permettere il raggiungimento di tale obiettivo. A tal proposito, si evidenzia che presso il SCC è già presente un sistema di raffreddamento basato su chilling ad acqua,

che dispone già della potenza frigorifera sufficiente per il raffreddamento, e che dovrà essere adeguato per il raffreddamento del nuovo sistema HPCC oggetto della presente fornitura.

Per la misura del PUE si dovranno considerare i seguenti valori ambientali:

- Temperatura Esterna Media=25°;
- Temperatura media dell'aria in aspirazione ai sistemi=25°;
- sistema di *chilling* completamente dedicato al raffreddamento del nuovo sistema HPCC.

Per il collaudo e le verifiche di raggiungimento del valore di PUE richiesto si provvederà a spegnere/disconnettere dal sistema di *chilling* esistente eventuali altre utenze che generino calore. Eventuali dati aggiuntivi necessari per la definizione ed il dimensionamento del sistema potranno essere acquisiti in fase di sopralluogo.

4 Infrastruttura rete dati

Il CMCC, nell'ambito del potenziamento e aggiornamento del suo data center, ha la necessità di potenziare il collegamento alla *backbone* di rete pubblica ed aggiornare l'infrastruttura di rete privata al fine di:

- migliorare e potenziare la connettività a livello geografico del SCC;
- aggiornare e potenziare la propria infrastruttura di rete privata;
- centralizzare le operazioni di management di tutti i sistemi (di calcolo, storage, networking ed altro) installati presso il data center del CMCC.

4.1 Core Datacenter Network (CDCN)

Il CMCC, nell'ambito del potenziamento e aggiornamento del suo data center, ha la necessità di aggiornare l'infrastruttura di rete privata seguendo i modelli e le principali soluzioni tecnologiche di ultima generazione.

Nei moderni data center, dal punto di vista del networking, risulta ormai superata la tradizionale architettura a tre livelli costituita dai livelli *access*, *aggregation* e *core*. Il motivo principale di ciò è da attribuire alla crescita del traffico di rete orizzontale (ovvero "east-west") all'interno del data center (server-server, server-storage, ecc.). Pertanto l'architettura di riferimento maggiormente in uso attualmente è quella denominata Clos-based (*leaf-spine*). Tale nuova architettura è stata progettata per minimizzare il numero di hops tra gli hosts.
Come si evince dalla figura Figura 1- Architettura leaf-spine, questo design appiattisce la topologia fisica, garantisce un'elevata scalabilità e fornisce una latenza predicibile switch-to-switch, rimuovendo quasi del tutto il rischio di loop di rete.

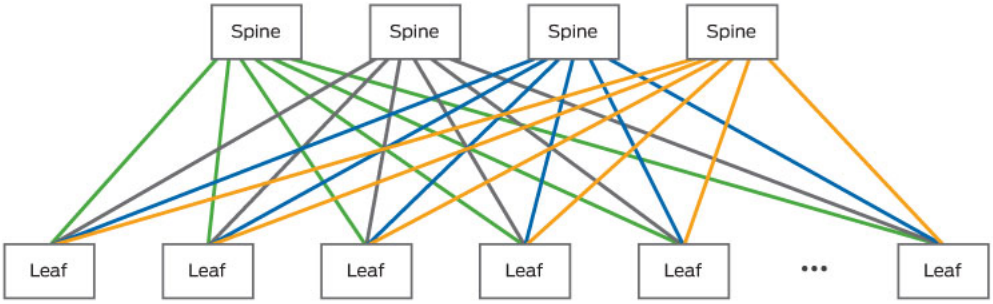


Figura 1- Architettura leaf-spine

Come illustrato nell'allegato A ("CMCC Supercomputing Center – Schema Generale"), l'infrastruttura di rete privata (di seguito denominata Core Datacenter Network "CDCN"), dovrà essere basata su topologia Clos-based e dovrà consentire l'interconnessione di sistemi eterogenei, dal cluster HPCC (almeno il 10% del numero dei nodi di calcolo e tutti i nodi di servizio del sistema HPCC), agli storage T2SS e DAS, ai servizi di data delivery e *identity management* a ogni altra entità di servizio del Supercomputing center.

La soluzione proposta dovrà avere tutte le caratteristiche generali ed i requisiti descritti nei paragrafi seguenti.

4.1.1 Caratteristiche Generali dell'architettura

L'architettura di connettività proposta dovrà implementare una fabric IP-Ethernet basata su protocolli standard ed open che dovrà apparire e comportarsi verso il mondo esterno come un unico sistema logico (Virtual Fabric). Dovrà essere pertanto possibile gestire, configurare, e automatizzare l'intera fabric come un singolo sistema logico.
La componente software che implementa le funzionalità richieste dovrà poter essere eseguita direttamente all'interno degli switch fisici, senza necessità di ricorrere a controller o altre entità esterne. L'insieme degli switch fisici richiesti dal presente Capitolato Tecnico e della componente software relativa all'implementazione delle funzionalità richieste dovranno pertanto essere autoconsistenti.
Il Sistema proposto dovrà implementare una fabric a singolo management plane, sia a livello locale che geografico, dotato di caratteristiche di alta affidabilità.
Dovrà essere possibile implementare la fabric su qualsiasi topologia fisica sottostante ed indipendentemente dall'interconnessione attraverso altri dispositivi non dello stesso vendor, senza la necessità che i sistemi siano adiacenti.
La Fabric dovrà poter essere gestita mediante CLI e RESTful API. La gestione mediante CLI/API dovrà supportare l'utilizzo dei più comuni tool di automazione, come ad es. Ansible o Python.
La fabric dovrà poter essere configurata come una singola entità, ed ogni switch appartenente alla fabric dovrà poter sincronizzare il suo provisioning state in maniera autonoma, con la possibilità di effettuare rollback a stati precedenti.
La Fabric dovrà permettere di configurare oggetti a livello sia globale che locale (singolo switch) e di identificare set di oggetti fabric-wide.

La fabric dovrà supportare eventuali situazioni di split-brain senza che ciò impatti sul forwarding del traffico locale. In caso di fabric-stretch su location geografiche differenti, il sistema dovrà poter continuare a mantenere validi e operativi sia il control plane che il data plane locali.

4.1.2 Gestione del Traffico

La fabric dovrà supportare meccanismi di:

- Broadcast suppression;
- Conversational forwarding;
- ARP Optimization (possibilità di effettuare proxy ARP se l'informazione è già presente nel DB interno);
- Anycast Gateway (possibilità per gli endpoint di utilizzare lo stesso virtual MAC/Indirizzo IP su tutti i first-hop switch). Le subnet interessate da funzionalità di anycast gateway dovranno poter essere configurate come un singolo oggetto atomico fabric-wide. Il numero di istanze VRF con capacità di anycast gateway dovrà essere pari almeno a 1000;

Le ottimizzazioni di forwarding sopra menzionate dovranno essere disponibili almeno con riferimento alle seguenti operazioni: bridging, routing, extended bridging (su VxLAN tunnels) ed extended routing (su VxLAN tunnels).
La fabric dovrà essere in grado di aggregare link tra due switch mediante meccanismi di Layer2 Multi-pathing e multi chassis/virtual chassis LAG.
Tutti gli switch nella fabric dovranno potersi scambiare informazioni topologiche relative ai device adiacenti e dovranno implementare uno shared endpoint database.

La fabric dovrà poter implementare:

- meccanismi STP standard con altri device non appartenenti alla fabric, utilizzando protocolli standard;
- un meccanismo di loop-guard di livello 2 non basato su STP. Dovrà cioè essere possibile, senza implementare meccanismi di STP con dispositivi esterni alla fabric, individuare la presenza di loop tra due qualsiasi porte appartenenti alla fabric anche quando le porte sono ubicate su switch differenti interconnessi localmente o attraverso una rete IP.

Gli switch forniti dovranno supportare la possibilità di incapsulare e decapsulare in HW il traffico VxLAN mediante meccanismi di VxLAN Tunnel EndPoint (VTEP).
La virtual fabric nel complesso dovrà supportare funzionalità di VTEP high availability (VTEP HA) attivando funzionalità di VTEP su almeno due switch: la VTEP HA dovrà poter usare lo stesso anycast IP address su entrambi gli switch, ed entrambi gli switch dovranno poter incapsulare e decapsulare il traffico VxLAN allo stesso tempo. La funzionalità di VTEP HA dovrà essere configurabile come un singolo fabric object (non dovrà richiedere l'effettuazione della configurazione su ogni singolo switch).
La fabric dovrà essere dotata di capacità native di interscambio di informazioni con altri nodi all'interno della fabric. In caso di VXLAN, queste informazioni dovranno includere almeno:

- i dati di configurazione del VXLAN Tunnel Endpoint (VTEP), ad esempio i VTEP IP address e le Virtual Network Interface (VNIs);

- le informazioni di raggiungibilità e di routing come i MAC Address, i VTEP IP address, e le informazioni VNI.

La fabric dovrà poter creare tunnel VxLAN verso altri nodi nella fabric senza utilizzare protocolli basati su multicast.
La fabric dovrà poter supportare all'interno di una istanza isolata di Sistema Operativo (p.es in un container o in una VM) l'implementazione di:

- Routing application stacks (OSPF o BGP);
- Interfacce OVSDB;
- Istanze virtual network manager.

I servizi implementati nelle istanze di Sistema Operativo dovranno poter essere portabili, ovvero migrabili verso nodi differenti della fabric.

La fabric dovrà inoltre supportare a livello fabric-wide funzionalità di:

- Policy-based Routing;
- Access Control;
- Line-rate control, manipolazione e redirezione logica o fisica dei flussi;
- Flow-level Security;
- Control Plane Traffic Protection: il traffic di control plane dovrà essere separabile e gestito con priorità rispetto al traffico di rete. Dovranno essere inoltre implementabili politiche di rate limiting al fine di mitigare gli effetti di eventuali attacchi di sicurezza.

4.1.3 Funzionalità di Datacenter Interconnect

La fabric dovrà poter estendere eventuali domini Layer 2 su più location geografiche remote, utilizzando protocolli IP standard come VxLAN al fine di garantire l'interoperabilità con eventuali reti e sistemi di terze parti. Non saranno ammesse soluzioni che facciano uso di protocolli o sistemi specializzati per il Data Center Interconnect.
Dovrà essere possibile realizzare tunnel DC verso altre location a partire da qualsiasi switch nella fabric.
La fabric dovrà poter estendere l'infrastruttura iperconvergente oggetto di acquisizione tra più data center, e dovrà poter raccogliere, organizzare e presentare mediante interfaccia grafica e testuale tutti i dati relativi alle connessioni (a titolo esemplificativo e non esaustivo: end-to-end latency, duration, total bytes transferred, stato delle connessioni TCP in real time) senza necessità di usare algoritmi di sampling e senza fare uso di wire taps o mirror ports.
In uno scenario multi-tenant, dovrà essere possibile implementare almeno i seguenti servizi:

- Point-to-Point Virtual Connections with HA;
- Multipoint-to-Multipoint Virtual Connections with HA;
- Fully transparent point-to-point pseudo-wires, inclusi i meccanismi di link state tracking e le informazioni di application telemetry.

La fabric dovrà supportare meccanismi avanzati di classificazione del traffico e di QoS e la capacità di regolare il traffico sia a livello di flusso che di porta fisica.

Dovrà essere possibile la tracciatura di qualsiasi flusso di traffico al fine di poter effettuare troubleshooting.

4.1.4 Virtual Networks / Multi-Tenancy

La fabric dovrà poter organizzare la rete fisica in più reti logiche (Virtual Networks o Tenant) distinte, ognuna dotata delle proprie risorse, servizi di rete e politiche QoS.
Ogni Tenant dovrà avere un singolo punto di management dedicato. Il Fabric Administrator dovrà poter assegnare l'ownership di ogni singolo tenant ad un amministratore distinto e dotato di credenziali separate, il quale potrà provvedere in autonomia alla gestione e configurazione del Tenant.
Ogni Tenant dovrà avere sia data plane che control plane isolati e separati.

4.1.5 Application Flows e Statistiche

La fabric dovrà implementare meccanismi di Application flow visibility in grado di fornire informazioni su:

- Real-time Correlation tra server e client;
- overlay/underlay Performance analysis;
- Capacity Planning & Network Troubleshooting sia a livello Intra-data center che Inter-data center;
- Security and Incident Response.

Gli switch dovranno supportare tool di sampling open-source come sFLOW e processi di metering come IPFIX in grado di fornire informazioni di traffico costanti su tutte le interfacce abilitate.

4.1.6 Management e monitoraggio

Dovrà essere reso disponibile ed incluso in fornitura un tool grafico di Fabric Management al fine di poter effettuare attività di Network Provisioning, Monitoring and Analytics.
Il tool di management dovrà poter visualizzare a livello grafico la completa topologia della rete, rappresentando tutti gli elementi che sono parte della fabric e gli eventuali device esterni ad essa interconnessi.
Attraverso il tool dovrà essere possibile effettuare il provisioning e il monitoraggio di multiple fabric instance, e dovrà altresì essere possibile automatizzare le operazioni relative al deployment iniziale della fabric.
Il tool di management dovrà almeno essere in grado di effettuare il provisioning ed il monitoraggio dei seguenti costrutti:

- Port characteristics;
- VLANs and port VLAN membership;
- STP characteristics;
- VLAN Interfaces and layer-3 ports;
- OSPF, BGP adjacencies and static routes;
- VxLAN VTEPs and related VxLAN to VLAN mappings;
- Point-to-Point transparent pseudowires;

- Security Policies;
- QoS Policies;
- Mirror objects.

Dovrà essere possibile raccogliere e visualizzare almeno i seguenti dati:

- fabric device model, serial number, transceiver inventory, software version, licensing information;
- fabric device system health, (CPU utilization, disk and memory utilization);
- Physical port counters over time, ingress and egress, con la possibilità di comparare fino a 10 porte nello stesso tempo;
- Endpoint active on each fabric device and related flow information;
- Syslog data;
- SNMP trap.

Il tool dovrà includere una funzionalità di network analytics collector in grado di raccogliere, analizzare e presentare all'utente i dati relative ai flussi applicative di rete, utilizzando almeno le seguenti sorgenti:

- Dati di telemetria della fabric;
- probe non-intrusivi Virtual Machine based al fine di consentire la visibilità di flussi di comunicazione inter-vm sullo stesso server;
- Altre sorgenti dati come ad es. Netflow, sFlow devono essere supportate.

I dati raccolti dovranno essere presentati in maniera chiara ed utile all'utente mediante grafici a torta, ad istogramma, diagrammi Sankey, ecc anche al fine di consentire l'identificazione di relazioni tra le applicazioni ed il funzionamento dell'infrastruttura fisica e di eventuali colli di bottiglia.

Dovrà essere possibile configurare:

- tag personalizzati che dovranno essere memorizzati insieme ai dati al fine di migliorare l'interpretabilità degli stessi da parte degli utenti;
- Nomi per specifici valori di porte TCP.

Dovrà essere possibile integrare il collector database con metadati provenienti da sistemi di terze parti come: server DNS, sistemi di geolocalizzaizone, sistemi di virtualizzazione/iperconvergenza (VMWare, Nutanix, ecc), Active Directory, ecc.
I dati raccolti dovranno poter essere filtrati e ricercabili mediante query specifiche impostate dall'utente.
Il tool dovrà mettere a disposizione una funzione di packet analytics che consenta la visibilità a livello L4-L7 dei pacchetti dati applicativi utilizzando almeno le seguenti sorgenti di dati:

- file PCAP offline;
- Traffico real-time.

4.1.7 Caratteristiche Hardware dell'Infrastruttura

Dovrà essere fornita una infrastruttura di *switching* ad alte prestazioni di classe data center, avente le seguenti caratteristiche HW minime:

- Architettura HW basata su chipset standard-silicon;
- Topologia fisica di tipo Spine-Leaf, con rapporto di over subscription non superiore a 3:1 e velocità di accesso alla rete pari ad almeno 10Gbps;
- L'infrastruttura dovrà mettere a disposizione almeno 300 porte di accesso di tipologia SFP+ o SFP28. Per il soddisfacimento del requisito, non sono ammesse eventuali porte oggetto di breakout (ad es porte a 100Gb splittate in più porte a 10 o 25Gb);
- Velocità di interconnessione tra i layer spine e leaf non inferiore a 100Gbps
- Alimentatori e ventole ridondati e hot-swap, assenza di ulteriori *single point of failure*.
- Supporto ONIE per l'utilizzo di sistemi operativi alternativi *linux-based*;
- Supporto di Sistemi Operativi di Rete differenti da quelli sviluppati dal produttore degli apparati. L'utilizzo di S.O. diversi non deve inficiare il supporto HW degli apparati;
- Dimensione massima di ogni switch max 2 RU;

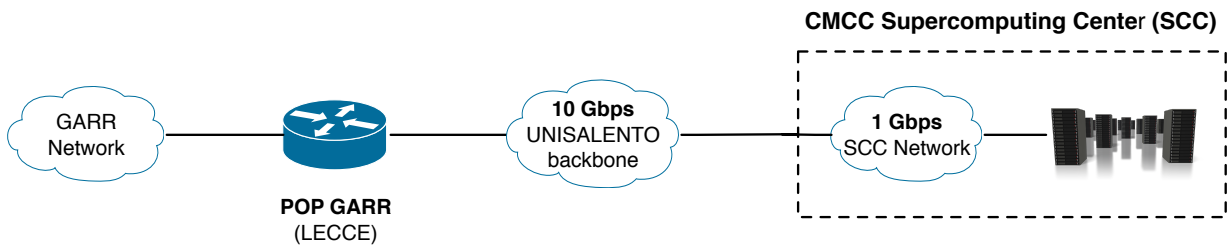
- Caratteristiche HW minime degli Switch di Accesso (Leaf):
  - Almeno 48 porte di accesso ad almeno 10Gbps con connettori SFP+;
  - Almeno 6 porte di uplink ciascuna a velocità di 100 Gbps QSFP28;
  - Switching capacity minima 1.6Tbps, non blocking;
  - Gli switch dovranno essere completi di ottiche 10Gb SR. Per l'interconnessione dei soli nodi oggetto della presente fornitura sarà ammesso l'uso di cavi DAC in rame esclusivamente per le connessioni all'interno dello stesso rack.

- Caratteristiche HW minime degli Switch di Core (Spine):
  - Almeno 32 porte 100Gbps QSFP28 con supporto delle velocità 10/25/40/50 Gbps;
  - Forwarding capacity minima: 4400 Mpps (Full Duplex, packet size >350bytes);
  - Switching capacity minima 6Tbps, non blocking.

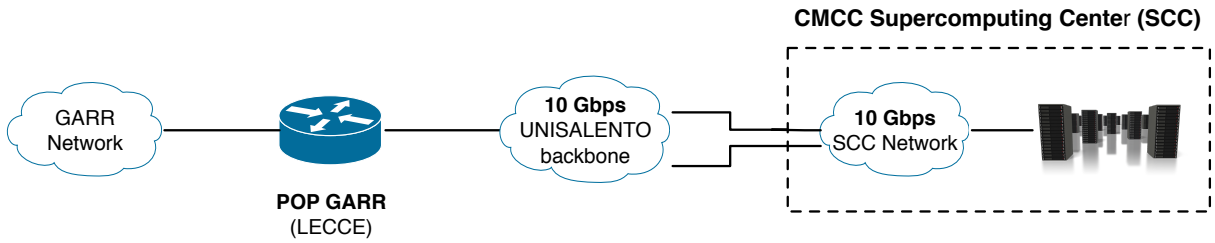
Le connessioni tra gli switch spine e leaf dovranno essere effettuate mediante cavi in fibra ottica. Fanno parte della fornitura le eventuali ottiche e cavi necessari per realizzare tutte le interconnessioni tra spine e leaf e verso tutti i sistemi oggetto della presente fornitura.

4.2 Data Center Backbone Connectivity (DCBC)

Il data center del CMCC, è situato all'interno del campus universitario Ecotekne dell'università del Salento. Attualmente, l'università del Salento è connessa alla rete GARR per mezzo di link a 10Gb/s mentre il data center del CMCC è connesso alla dorsale di rete Unisalento con un link in fibra a 1Gb/s come descritto nella seguente figura.



Al fine di potenziare il collegamento geografico del SCC, il CMCC intende attivare due nuovi link in fibra a 10Gb/s tra SCC e la dorsale di rete Unisalento come evidenziato nella figura di seguito.



Si richiede pertanto la fornitura, all'interno della infrastruttura di rete CDCN, delle ottiche o degli eventuali altri apparati necessari per realizzare ed attivare i due nuovi collegamenti a 10Gb/s sopra descritti.

4.3 Management Data Centre Network (MDCN)

Al fine di rendere possibile il management di tutti i dispositivi oggetto del seguente capitolato, si richiede la fornitura di una rete di management esterna (out-of-band) ovvero una rete con link dedicato e separato dai link principali.
La tipologia di rete richiesta è Ethernet con velocità di almeno di 1Gbps.
La fornitura dovrà prevedere un certo numero di switch Layer 2 con un opportuno numero di porte tali da ospitare i collegamenti per le interfacce di management e le interfacce tipo IPMI/IMM/IDRAC ecc. La fornitura dovrà essere compresa di opportuno cablaggio.

5 Sistemi e servizi di storage

La necessità di gestire il ciclo di vita dei dati e delle informazioni in essi contenute rende sempre più evidente il bisogno di catalogare e memorizzare tali dati all'interno di una infrastruttura di storage multi-tier. Tale infrastruttura deve poter consentire il soddisfacimento di differenti esigenze di archiviazione a breve, medio e lungo termine garantendo, nello stesso tempo, adeguati livelli di disponibilità e reperibilità dei dati.

E' quindi indispensabile poter disporre di sistemi di storage (di diversa natura rispetto agli storage HPC solitamente annessi ai sistemi di calcolo) destinati ad accogliere i dati da disseminare tramite opportuni servizi dedicati ed i dati consolidati, acceduti meno frequentemente. E' altrettanto opportuno poter definire ed utilizzare storage pool dedicati con caratteristiche logiche (tempi di retention dei dati) e fisiche (tempi di accesso al dato, throughput, capacità, ecc) differenti.

Pertanto, si richiede la fornitura di:

- uno storage di tipo Tier2 (T2SS), di tipo file-based, per accogliere:
  - le home directory degli utenti dell'SCC;
  - i dati prodotti dalle simulazioni numeriche e destinati ad essere disseminati tramite servizi di data delivery;
  - in generale, tutti quei dati che è necessario mantenere online.

- un Deep Archive Storage System (DASS), di tipo object-based, per l'archiviazione dei dati nel medio e lungo termine.

5.1 Tier2 Storage System (T2SS)

Il sistema di storage T2SS dovrà essere un'infrastruttura storage condivisa per la gestione del data lake, orientata alla fornitura di servizi, con un elevato grado d'indipendenza dalle infrastrutture di elaborazione dati e dalle applicazioni connesse, accessibile da tutti i sistemi interconnessi alla rete Core Datacenter Network "CDCN" e con un buon compromesso tra performance e capacità di memorizzazione dati.

Il sistema di storage utilizzato per il T2SS dovrà essere di livello 2, ovvero TIER2. Generalmente lo storage TIER2 è considerato come storage "online", appartenente alla classe di storage che soddisfa i requisiti di grande capacità di archiviazione, accesso ai dati istantaneo, buoni livelli di scalabilità e protezione.

Il sistema proposto, completo di sottosistema hardware e software, dovrà essere una soluzione per la gestione di dati non strutturati, ad accesso file level mediante servizi erogati attraverso protocolli IP. In particolare il sistema dovrà avere le seguenti caratteristiche:

- L'architettura proposta dovrà essere di tipo Scale-out Network Attached Storage (NAS);
- La capacità utile del sistema dovrà essere di almeno 2 PebiBytes;
- Le prestazioni di I/O in lettura dovranno essere di almeno 5 GBytes/s;
- Le prestazioni di I/O in scrittura dovranno essere di almeno 3 GBytes/s;
- Il sistema dovrà essere capace di espandere le performance e la capacità linearmente. Gli incrementi di performance e capacità storage lineari dovranno essere effettuabili aggiungendo nodi storage, ciascuno con i suoi dischi, cache, I/O e potenza computazionale (CPU) per assicurare la scalabilità lineare;
- Tutti i nodi storage/controller dovranno essere attivi e contribuire alle performance e alla capacità del sistema;
- Il sistema storage dovrà consentire l'espansione con hardware di nuova generazione, senza cambiamenti alla configurazione esistente e mentre il sistema è online. Dovrà consentire inoltre la dismissione (eviction) di hardware di vecchia generazione se e quando richiesto;
- L'architettura dovrà supportare il bilanciamento automatico e senza interruzione del servizio dei dati per ottenere performance ottimali e efficienza della capacità in caso di espansioni successive del sistema;
- L'accesso dei client al file system e alle share dovrà essere automaticamente distribuito su tutti gli chassis/nodi;
- Il Sistema dovrà avere una cache coerente globale, scalabile quando vengono aggiunti più nodi al cluster;
- Il sistema di storage dovrà supportare l'accesso ai dati tramite collegamento Ethernet sia a 10Gbps che a 40Gbps sul frontend.
- Il sistema dovrà fornire l'accesso per una varietà di sistemi operativi (UNIX, MAC, Linux, Windows) usando tutti i protocolli standard: NFSv3, NFSv4, NFS Kerberized sessions (UDP o TCP),

SMB1 (CIFS), SMB2, SMB3, SMB3-CA, SWIFT, HTTP, FTP, NDMP, SNMP, LDAP. Tutti i protocolli dovranno essere inclusi senza licenze addizionali o ulteriore hardware;

- Il sistema dovrà supportare l'autenticazione degli utenti e degli amministratori con NIS, LDAP e Active Directory;
- Il Sistema dovrà essere dotato di funzionalità di Journal File System;
- Il sistema dovrà consentire di creare differenti storage pool di capacità e performance composti da dischi di tipologia differente (SAS, SATA e SSD) con un file system unico, fornendo agli utenti finali e alle applicazioni capacità aggregata e la visione delle performance del sistema;
- Il sistema dovrà implementare un singolo file system, scalabile fino a più di 50PB, scalabile dinamicamente in un singolo namespace per l'intera capacità fornita, senza strati di virtualizzazione e senza la necessità di installare software addizionale sui client;
- Il sistema dovrà essere in grado di gestire il ciclo di vita dei dati e migrare i file tra i differenti storage pool, utilizzando politiche basate sull'età del file, sul tipo, sulla dimensione e sulla posizione nelle directory;
- Il file system dovrà essere continuamente e automaticamente bilanciato su tutti i nodi e i dischi, per eliminare colli di bottiglia e zone calde;
- Il file system dovrà permettere un numero illimitato di accessi client indipendentemente dal sistema operativo e dal protocollo;
- Il sistema dovrà supportare l'impostazione di quote utenti con limiti soft o hard ed Over Provisioning;
- Il sistema dovrà supportare API Restful Access to Namespace (RAN) per l'integrazione di applicazioni esterne per funzionalità di file sync e share;
- Il sistema di storage dovrà fornire un modo per bilanciare le performance di accesso ai dati verso tutti i client sulla rete ed inoltre dovrà essere in grado di allocare ampiezze di banda differenti ai differenti client, al fine di soddisfare differenti esigenze di performance;
- Il sistema dovrà supportare la replica almeno asincrona;
- Il sistema dovrà fornire protezione garantita da scenari di corruzione dati silenti;
- Il sistema di storage dovrà essere in grado di sostenere fallimenti di dischi multipli e di controller multipli;
- Il sistema di storage dovrà restare completamente online e con tuti i dati accessibili anche in caso di un fallimento di un intero enclosure/nodo;
- Il sistema dovrà consentire di cambiare il livello di protezione in maniera granulare a livello di sistema, directory o file.
- dovrà essere possibile eseguire la creazione, il restore e la cancellazione automatica di snapshot;
- Il sistema dovrà supportare il Reporting avanzato e l'analisi delle performance, del trend dello storage e strumenti di capacity planninng;
- Il sistema dovrà essere dotato di interfaccia Web e CLI;
- Il sistema dovrà supportare l'SNMP monitoring;
- Il sistema dovrà consentire il monitoraggio della capacità ed il reporting a livello di directory, utenti e gruppi;
- Il sistema dovrà supportare lo storico delle performance e la loro analisi;
- Il sistema dovrà fornire funzionalità di monitoraggio remoto e di "call home" al fine di allertare automaticamente il centro di supporto in caso di eventuali guasti e/o richieste di manutenzione.

5.2 Deep Archive Storage System (DASS)

Al fine di gestire l'archiviazione dei dati nel medio e lungo termine, si richiede la fornitura di un Deep Archive Storage System (DASS), di tipo object-based.

Il sistema DASS dovrà essere dotato di tutti gli apparati hardware e dei relativi pacchetti software di tipo enterprise necessari per consentire l'accesso alle risorse di storage del DASS (tape library, storage nearline, ecc…), per gestire i servizi di archiviazione a medio e lungo termine, per agevolare le interazioni con le altre tipologie di storage (sia quelle richieste nel presente capitolato tecnico che quelle future) e per favorire l'integrazione veloce e semplificata con software di Data and Metadata Management di terze parti (open-source e non).
Inoltre, per mezzo del relativo front-end d'accesso e gestione, il sistema dovrà essere in grado di contattare varie tipologie di storage, on-site ed off-site, tramite protocollo S3 e di garantire l'alta disponibilità di tutti i servizi correlati.
Il front-end dovrà essere connesso alla Core Datacenter Network CDCN ad almeno 10Gbps al fine di garantire un throughput pari ad almeno 1,25GBytes/sec in fase di scrittura e lettura di dati verso il DASS. Infine, tale throughput dovrà essere in ogni caso garantito anche a fronte di un importante intervento di potenziamento degli storage da esso gestiti ed acceduti.

Per l'implementazione del DASS si richiede inoltre la fornitura dei dispositivi hardware e dei sistemi software di seguito descritti.

5.2.1 Tape Library

La tape library è destinata ad accogliere i dati consolidati ed a memorizzarli su nastro a medio e lungo termine. In particolare, la tape library dovrà soddisfare le seguenti caratteristiche/funzionalità:

- capacità iniziale pari ad almeno 2 Pebi Bytes di spazio utile (dovrà essere espandibile fino ad almeno 12 Pebi Bytes);
- utilizzo di frame ad alta densità;
- n. 6 tape drive di nuova generazione basati su tecnologia LTO8 (o superiore) con interfaccia Fibre Channel ad almeno 8Gbps;
- throughput nativo del singolo drive uguale (o superiore) a 360 MB/s;
- support ALMS (Advanced Library Management System) per la gestione dinamica dello storage;
- funzione di autocalibrazione;
- gestione locale e remota della tape library e delle risorse in essa contenute;
- funzionalità di media lifecycle management;
- funzionalità di drive lifecycle management;
- funzionalità di library lifecycle management;
- auto drive clean;
- supporto partizioni multiple all'interno della singola tape library;
- ottimizzazione dell'utilizzo delle risorse;
- espandibilità della capacità di storage senza alcun impatto sui dati già memorizzati;
- interfaccia grafica (GUI) per la gestione della tape library;
- supporto di connettività eterogenea/multivendor;

- sicurezza ed integrità del trasferimento dati;
- alta affidabilità;
- facilità di gestione;
- possibilità di aggiungere (a posteriori) nuove frame e nuovi drive di uguale (o superiore) tecnologia;
- fornitura di nastri scratch LTO8 (o superiore), completi di bar-code label, in quantità tale da soddisfare la capacità utile minima precedentemente indicata (2PiB).

Ulteriori caratteristiche della tape library:

- automatic control-path;
- automatic data-path;
- tape-drive compression ed encryption;
- accessori robotici ridondati (almeno due);
- archiviazione onsite ed offsite;
- supporto per la gestione dinamica dello storage, delle librerie logiche e della configurazione dei drive;
- verifica integrità dei media automatica.

5.2.2 Storage Area Network

E' prevista la fornitura di una SAN di nuova generazione ad almeno 8Gbps ridondata, composta da almeno 2 switch Fibre Channel con il 50% di porte libere, per collegare i diversi elementi che costituiscono il DASS (front-end del sistema DASS ed i tape drive).
Dovrà essere prevista la topologia switched fabric (FC-SW) con zoning per la tecnologia SAN.

5.2.3 Monitoraggio e gestione del DASS

Il DASS dovrà, inoltre, essere dotato di un sistema di monitoraggio e gestione centralizzato.
Tali sistemi dovranno consentire di:

- valutare lo stato di salute di ognuna delle componenti del DASS;
- segnalare tempestivamente eventuali situazioni di failure;
- effettuare eventuali attività di manutenzione sui dispositivi in failure;
- effettuare eventuali attività di aggiornamento del software e del firmware utilizzato da ciascuna delle componenti del DASS.
- effettuare analisi statistiche sui dati di utilizzo delle risorse, ecc;
- gestire, tramite un sistema di quote, lo spazio di archiviazione associato ad ogni utente o gruppo di utenti utilizzatori.

6 Adeguamento dell'infrastruttura impiantistica

Al fine di garantire il corretto funzionamento dei sistemi che verranno forniti, il fornitore dovrà farsi carico di tutti gli eventuali interventi di adeguamento degli impianti tecnologici presenti presso il SCC. In particolare, il fornitore prima di avviare qualsiasi attività di installazione dei sistemi, dovrà provvedere a:

- eventuali interventi di adeguamento sull'impianto elettrico esistente;
- eventuali interventi di adeguamento sull'impianto di condizionamento esistente.

Tali eventuali interventi dovranno, in ogni caso, essere concordati preliminarmente con il Responsabile Tecnico del SCC.
Durante il periodo che intercorre tra la pubblicazione del bando ed il termine di presentazione delle offerte, gli interessati potranno chiedere informazioni sull'attuale infrastruttura impiantistica.
Inoltre, a tal proposito, è richiesto al fornitore, pena esclusione dal procedimento di gara, un sopralluogo obbligatorio, prima della presentazione dell'offerta, dei luoghi, degli impianti e di quant'altro necessario per una corretta installazione ed un efficiente funzionamento dei sistemi che saranno forniti.

7 Servizi di supporto

A supporto della fornitura oggetto del presente bando, si richiedono i seguenti servizi:

- Servizio di formazione e supporto per una corretta conduzione operativa dei sistemi e dei servizi forniti;

- Servizio di manutenzione hardware e software per tutti sistemi forniti.

7.1 Servizio di Formazione e Supporto per una corretta conduzione operativa dei servizi e dei sistemi

Il "Servizio di Formazione e Supporto" dovrà essere rivolto al personale tecnico del Supercomputing Center del CMCC che ha la responsabilità di gestire e mantenere operativa l'intera infrastruttura di calcolo e storage in dotazione al Centro. L'obiettivo di tale servizio sarà quello di fornire al personale tecnico incaricato la necessaria formazione ed il supporto per lo svolgimento delle seguenti attività:

- la definizione, la realizzazione e l'esecuzione delle procedure di gestione della fornitura;
- il mantenimento delle prestazioni della fornitura;
- il mantenimento e l'aggiornamento delle configurazioni hardware e software di tutta la fornitura.

Le attività necessarie all'espletamento del "Servizio di Formazione e Supporto" saranno di tipo "training on job" e saranno svolte da personale incaricato dal Fornitore e con competenze specialistiche adeguate alle attività richieste.

7.2 Servizio di manutenzione hardware e software

L'importo totale dell'appalto include anche il servizio di manutenzione hardware e software a copertura di tutti i componenti previsti nella fornitura. I requisiti per tale servizio sono i seguenti:

- servizio di manutenzione di tipo "*full service on site*" e Next Business Day;
- durata del servizio: il servizio decorrerà dalla data di accettazione della fornitura e avrà le seguenti distinte durate:
    - 3 anni per i soli nodi di calcolo del sistema HPCC;
    - 5 anni per la restante parte della fornitura.

Considerata la complessità della fornitura, si ritiene indispensabile che tale servizio sia erogato direttamente dai costruttori/produttori delle componenti hardware e software con i quali il personale tecnico del CMCC interagirà direttamente senza intermediazione del fornitore. Al fornitore sarà eventualmente demandata l'attività di sostituzione delle parti hardware dichiarate guaste dal produttore/costruttore.

A maggior chiarimento, il servizio di manutenzione e assistenza tecnica richiesto comprende:

- la fornitura degli aggiornamenti e delle revisioni (patch, minor e major release, ecc..) di tutto il software di base e applicativo in fornitura (sistema operativo, software vari, ecc..) nonché del firmware. In particolare, qualora il software fornito fosse sostituito con software equivalente e/o con potenzialità superiori, commercializzato con lo stesso nome o con nomi differenti da quello con cui è stato inizialmente fornito, il CMCC potrà richiederlo a costo zero e alle stesse condizioni di licensing;
- consulenza telefonica specialistica sul software di base e applicativo (sistema operativo, software terze parti, tuning, ecc..);
- la predisposizione e mantenimento del Piano di Recovery;
- la sostituzione, presso la sede del SCC, di tutti i componenti guasti senza alcun onere aggiuntivo per il CMCC (come ad esempio, costi di manodopera, di spedizione, di trasferte, ecc.);
- la presa in carico del malfunzionamento entro un tempo massimo di 4 ore dalla segnalazione (alla quale dovrà essere associato il relativo trouble ticket);
- l'intervento per la risoluzione dei malfunzionamenti hardware e/o software nel rispetto della tempistica riportata nella tabella seguente:

Tabella 1

Livello di Gravità	**Definizione**	**Tempi di risoluzione**
Livello 1 – Alto Impatto	Con riferimento ai sistemi oggetto della fornitura, si verifica l'indisponibilità totale di almeno uno di essi.	Entro il tempo max di 48 ore solari
Livello 2 – Medio Impatto	Con riferimento ai sistemi oggetto della fornitura, si verifica la parziale indisponibilità di almeno uno di essi.	Entro il tempo max di 96 ore solari
Livello 3 – Basso Impatto	Si verifica un fault, su uno qualsiasi dei sistemi oggetto della fornitura, che non ne pregiudica il corretto funzionamento.	Entro il tempo max da concordare ma comunque non superiore a 168 ore solari.

Il tempo di risoluzione della malfunzione riportato nella suddetta tabella è da intendersi a partire dalla data e ora di segnalazione della malfunzione. Per risoluzione è da intendersi il ripristino delle condizioni di funzionamento e delle configurazioni esistenti prima dell'avvenuto guasto/malfunzionamento hardware e/o software.

Le attività necessarie alla risoluzione della malfunzione, a discrezione del CMCC, potranno proseguire ad oltranza anche nelle giornate di sabato, di domenica e/o in giorni festivi.
La malfunzione terminerà con la risoluzione del problema.

8 Modalità di fornitura e installazione

Il servizio di consegna ed installazione presso la sala computer del SCC dovrà essere erogato dal Fornitore, attraverso personale specializzato. A tal proposito è richiesto al fornitore un sopralluogo obbligatorio prima della presentazione dell'offerta, pena esclusione dal procedimento di gara, dei luoghi, degli impianti e di quant'altro necessario per una corretta installazione ed un efficiente funzionamento dei sistemi offerti.
Tutte le attività si intendono comprensive di ogni onere relativo al trasporto, facchinaggio, consegna "al piano", posa in opera, asporto dell'imballaggio e di qualsiasi altra attività ad esse strumentale.
Il Fornitore, inoltre, dovrà dotarsi di mezzi opportuni e/o di quanto altro necessario a trasportare, scaricare e a collocare la fornitura nella sala suddetta.
Il Fornitore garantirà, durante tutte le fasi di lavorazione, il rispetto delle normative vigenti in materia di tutela della salute e della sicurezza nei luoghi di lavoro.

9 Collaudo e Accettazione

Durante il periodo di test, della durata massima di 30 (trenta) giorni solari a partire della data di sottoscrizione del "Verbale di consegna", personale tecnico del CMCC, coadiuvato da personale del Fornitore, provvederà alle seguenti verifiche mirate al collaudo.

9.1 Verifica del rispetto dei requisiti tecnici della fornitura

La verifica dovrà accertare che la fornitura, per quanto riguarda il numero e la tipologia dei componenti, tecniche e metodologie impiegate, l'esecuzione e le funzionalità, siano in tutto corrispondenti a quanto previsto dai documenti della procedura in questione.

9.2 Collaudo funzionale del sistema di calcolo scalare/parallelo (HPCC)

Al fine di verificare il corretto funzionamento del sistema ed il rispetto di tutte le funzionalità richieste, dovranno essere eseguite tutte le prove di seguito descritte:

- Esecuzione del benchmark STREAM: al fine di verificare il corretto funzionamento della memoria RAM e la relativa larghezza di banda, dovrà essere eseguito il benchmark STREAM su tutti i nodi del sistema HPCC offerto (sia nodi di calcolo che nodi di servizio).

Valore di accettazione per tale prova:

il raggiungimento di almeno l'80% della larghezza di banda della memoria dichiarata nelle specifiche tecniche dei produttori dell'hardware fornito.

- Esecuzione dei test di performance della rete Infiniband (Perftest): al fine di verificare il corretto funzionamento della rete d'interconnessione veloce intra cluster Infiniband e la relativa larghezza di banda e latenza, dovranno essere eseguiti i test *ib_write_bw, ib_write_lat, ib_read_bw, ib_read_lat* forniti con la suite "*Perftest*". Tali test, che coinvolgeranno coppie di nodi, dovranno essere effettuati in maniera tale da verificare le performance di tutti i nodi che compongono il sistema. I test dovranno essere eseguiti nelle seguenti condizioni:

  - ogni test dovrà essere eseguito singolarmente e non in concorrenza con altri test;
  - il test tra due nodi distinti dovrà essere eseguito in modalità bidirezionale (con l'opzione "-b");
  - ciascun test dovrà essere eseguito per una durata non inferiore a 20 sec.

  Valore di accettazione per tale prova:
  Tutti i test saranno superati se i valori di larghezza di banda e latenza ottenuti risulteranno pari ad almeno l'85% dei valori dichiarati dal costruttore.

- Esecuzione del benchmark HPL (High Performance Linpack): al fine di verificare il corretto funzionamento del sistema nel suo complesso e verificare la prestazione aggregata di calcolo, dovranno essere eseguiti due run distinti del benchmark nelle seguenti condizioni:

  - Dovranno essere coinvolti tutti i nodi di calcolo del sistema;
  - Dovrà essere utilizzata la rete di interconnessione Infiniband;
  - I run in questione dovranno essere sottomessi tramite il job scheduler che sarà fornito con il sistema;
  - I run dovranno avere una durata di almeno 10 ore.

  Valore di accettazione per tale prova:
  Il test sarà superato se entrambi i run raggiungeranno almeno il 60% della capacità di calcolo aggregata di picco (*theoretical peak performance*) offerta.

9.3 Collaudo funzionale del sistema di storage HPC (HPCSS)

Al fine di verificare il corretto funzionamento del sistema di storage HPCSS e la relativa performance di I/O, dovranno essere eseguite tutte le prove di seguito descritte:

- Esecuzione del benchmark IOR: l'obiettivo di tale prova è quello di misurare e verificare le performance aggregate di I/O sul cluster parallel file system che sarà implementato sul sistema di storage HPCSS. Per tali prove dovrà essere utilizzato il benchmark IOR (https://github.com/hpc/ior). I test dovranno essere eseguiti nelle seguenti condizioni:

  - Dovrà essere utilizzata la rete di interconnessione Infiniband;
  - Dovrà essere utilizzata l'interfaccia di accesso ai dati POSIX;
  - I run in questione dovranno essere sottomessi tramite il job scheduler che sarà fornito con il sistema;
  - Le dimensioni del block size e del file size aggregato dovranno essere tali da minimizzare l'effetto della cache e bilanciare il workload;

- Prima di ogni test, dovranno essere eseguiti su tutti i nodi le operazioni di cancellazione della cache ed il remount del file system.

Valore di accettazione per tale prova:
Il test sarà superato se in almeno 10 esecuzioni distinte del benchmark IOR si raggiungerà il valore di performance di I/O aggregato richiesto nel paragrafo 3.6. Le 10 esecuzioni distinte dovranno essere effettuate con numero variabile di thread e client nonché con dimensioni variabili di block size e file size al fine di dimostrare la scalabilità della soluzione offerta. Inoltre, almeno uno di tali test dovrà coinvolgere almeno il 50% del numero dei nodi di calcolo del sistema HPC ed un numero di thread pari ad almeno il doppio del numero di meccaniche fisiche presenti nella soluzione di data storage HPCSS offerta.

9.4 Accettazione della fornitura

Al termine delle suddette verifiche sarà redatto il "Verbale di collaudo" in contraddittorio con il Fornitore. Nel caso di esito positivo del collaudo la data del suddetto verbale sarà considerata quale "Data di accettazione fornitura". Nel caso di esito negativo il Fornitore dovrà eliminare, entro 30 giorni solari, i vizi accertati. Test e collaudo saranno ripetuti e ad ulteriore esito negativo saranno applicate le penali, fino alla rescissione del contratto, come previsto dagli ulteriori documenti di gara.

10 Modalità di presentazione dell'offerta tecnica

L'offerta tecnica dovrà essere presentata mediante una Relazione tecnica, preferibilmente non più lunga di 50 pagine, stilata seguendo l'indice e gli argomenti del presente capitolato tecnico.

11 Certificazioni e requisiti del fornitore

Il fornitore deve possedere certificazioni di conformità del proprio Sistema Qualità alla norma UNI EN ISO 9001 con settore e scopo della certificazione coerenti con l'oggetto dell'appalto, rilasciata da organismo accreditato. Gli ulteriori requisiti del fornitore sono descritti nel documento "Disciplinare di Gara".

12 Penali

Nella tabella di seguito allegata, si descrivono le penali previste relativamente alle attività di consegna, accettazione e manutenzione della fornitura:

Vincoli di fornitura	Parametri da rispettare	Tolleranze/ritardi ammessi	Penali applicate al superamento delle soglie consentite
Tempi di consegna e messa in servizio	Entro 75 giorni solari dall'emissione dell'ordine	-	Per ogni giorno solare di ritardo, 0,1% dell'importo del contratto relativo alla fornitura in consegna
Accettazione della fornitura	Entro 30 giorni dalla consegna e messa in servizio	Ulteriori max 30 giorni in caso di primo collaudo con esito negativo	Per ogni giorno solare di ritardo, 0,1% dell'importo del contratto relativo alla fornitura in consegna
Risoluzione delle malfunzioni hardware e/o software con livello di gravità 1 "**Alto impatto**"	Entro il tempo max di 48 solari dalla segnalazione	Solo per cause di forza maggiore documentate dal fornitore ed accettate dal CMCC	Per ogni giorno solare di ritardo, 0,2% dell'importo del contratto relativo alla fornitura in consegna
Risoluzione delle malfunzioni hardware e/o software con livello di gravità 2 "**Medio impatto**"	Entro il tempo max di 96 solari dalla segnalazione	Solo per cause di forza maggiore documentate dal fornitore ed accettate dal CMCC	Per ogni giorno solare di ritardo, 0,02% dell'importo del contratto relativo alla fornitura in consegna
Risoluzione delle malfunzioni hardware e/o software con livello di gravità 3 "**Basso impatto**"	Entro il tempo max di 168 solari dalla segnalazione	Solo per cause di forza maggiore documentate dal fornitore ed accettate dal CMCC	Per ogni giorno solare di ritardo, 0,02% dell'importo del contratto relativo alla fornitura in consegna

L'applicazione delle penali è ad insindacabile giudizio del CMCC.

13 Unità di misura

13.1 Spazio disco RAW

Come standard in ambito storage, le unità di misura delle capacità sono da intendersi su base decimale. Pertanto dove si fa riferimento a dimensioni multiple di Byte con riferimento a spazio disco fisico "RAW", sono valide le seguenti relazioni:

Kilobyte (KB) = 1.000 bytes
Megabyte (MB) = 1.000.000 bytes
Gigabyte (GB) = 1.000.000.000 bytes
Terabyte (TB) = 1.000.000.000.000 bytes
Petabyte (PB) = 1.000.000.000.000.000 bytes

13.2 Spazio disco utile

Come standard in ambito informatico, le unità di misura delle capacità sono da intendersi su base binaria. Pertanto dove si fa riferimento a dimensioni multiple di Byte con riferimento a spazio disco "utile", sono valide le seguenti relazioni:

Kibibyte (KiB) = 1.024 bytes
Mebibyte (MiB) = 1.048.576 bytes
Gibibyte (GiB) = 1.073.741.824 bytes
Tebibyte (TiB) = 1.099.511.627.776 bytes
Pebibyte (PiB) = 1.125.899.906.842.624 bytes

In particolare indicheremo “spazio disco utile”, lo spazio disco effettivamente utilizzabile da utenti ad applicazione, vale a dire quindi lo spazio disco al netto di tutte le componenti necessarie all’utilizzo (ad esempio, cache, sistemi di mantenimento della consistenza, ecc), che non rientrano nel conteggio.

Allegato A: Schema generale CMCC Supercomputing Center