

Il CASPUR sceglie l'infrastruttura EMC per gestire i 'Big Data' generati dai supercalcolatori

Le tecnologie EMC VPLEX, CLARiiON e l'ambiente VMware alla base della nuova architettura del consorzio internuniversitario per il supercalcolo: prestazioni e throughput elevato, flessibilità e affidabilità. Piattaforma per servizi di Cloud Computing



In breve

La sfida

Riorganizzare le risorse storage con un'infrastruttura flessibile e affidabile per supportare le applicazioni Data-Intensive dei supercalcolatori e per realizzare una piattaforma per i servizi di Cloud Computing.

La soluzione

Ambiente VMware
Storage EMC CLARiiON Cx4
EMC VPLEX

Benefici

- Maggiori prestazioni e disponibilità dati
- Grande flessibilità di ampliamento della capacità storage
- Elevata affidabilità
- Storage tiering per soddisfare le diverse esigenze delle applicazioni
- Sensibile diminuzione dei consumi energetici
- Piattaforma ideale per i servizi di Cloud Computing
- Predisposizione al disaster recovery

Il Consorzio interuniversitario per le Applicazioni di Supercalcolo Per Università e Ricerca (CASPUR, www.caspur.it) nasce nel 1992 per volontà del Ministero dell'Istruzione, Università e Ricerca con l'obiettivo di fornire le tecnologie più innovative e le migliori competenze alla comunità italiana della ricerca. In questi vent'anni il CASPUR ha convertito l'obiettivo in realtà mettendo a disposizione delle attuali 11 università consorziate e del Ministero uno tra i più significativi centri di high performance computing (HPC) e una serie di servizi avanzati per la Pubblica Amministrazione.

Il CASPUR, con sede a Roma, è il centro di supercalcolo più importante dell'Italia centro-meridionale dove lavorano circa 180 tra dipendenti e collaboratori estremamente qualificati, dai quali dipende una mole consistente e crescente di attività, riconducibile negli ultimi due anni a più di un milione di accessi riguardanti oltre 400 progetti nelle diverse aree delle scienze computazionali, per un consumo complessivo superiore a 30 milioni di ore di processo.

TECNOLOGIE E COMPETENZE

Oltre ad offrire ai clienti grandi potenze computazionali, che negli altri centri di supercalcolo rappresenta la funzione essenziale, il CASPUR si avvale di infrastrutture tecnologiche di avanguardia e di particolari competenze in ambito scientifico, tecnico, applicativo e gestionale. I circa 35 dipendenti della divisione HPC sono in possesso di skill specifici che si riferiscono a modelli matematici e di calcolo per fluido-dinamica, scienza dei materiali, chimica, biologia, bioinformatica, statistica e altre specializzazioni. Competenze che permettono di affiancare i clienti nei più avanzati ed impegnativi progetti di ricerca per raggiungere il miglior risultato nelle principali aree delle scienze computazionali.

Pioniere in Italia per le tecnologie di supercalcolo e di interconnessione, il CASPUR è stato il primo nel 2008 ad adottare sistemi basati su processori Opteron con il cluster 'Matrix', costituito da 384 high performance computing nodes in grado di erogare oltre 20 TFLOPS (triloni di operazioni al secondo).

Più recentemente è stato aggiunto il cluster 'Jazz', un sistema di calcolo ad alte prestazioni basato su acceleratori grafici Fermi NVIDIA (l'architettura di GPU Computing – unità di elaborazione grafica – più avanzata mai realizzata) che a parità di prestazioni riduce i consumi ad un decimo rispetto a 'Matrix'. Con la sua tecnologia, il cluster 'Jazz' si è distinto al primo posto in Europa e

al quinto nel mondo nella classifica 2010 di Little Green 500, caratterizzando il CASPUR come uno dei centri di supercalcolo più ecologicamente sostenibili del pianeta.

Il Consorzio ha, inoltre, fortemente investito nell'acquisizione di competenze sia sistemistiche sia applicative sulla nuova architettura di calcolo, al punto di essere recentemente nominato da NVIDIA *CUDA Research Center* (riconoscimento dato alle istituzioni che sostengono e utilizzano il GPU Computing in molteplici campi di ricerca), attributo che lo pone in primo piano in Italia nello sviluppo e nel supporto al porting di applicazioni su architetture many-core.

Altro importante ruolo del CASPUR è quello di fornire servizi Ict alla Pubblica Amministrazione e ad altri enti terzi che necessitano di supporto per progetti applicativi. L'area servizi si basa su blade systems virtualizzati in ambiente VMware sui quali sono attive oltre 200 macchine virtuali che, oltre ai benefici in termini di gestione e flessibilità, hanno permesso di ridurre di circa un terzo i consumi rispetto alla situazione precedente.

PROCESSI SEMPRE PIÙ DATA-INTENSIVE

“Fino a pochi anni fa, un centro HPC si caratterizzava in base alla potenza di calcolo che poteva ad erogare” commenta Nico Sanna della struttura HPC del CASPUR. *“Oggi il panorama è sostanzialmente cambiato. Nella nostra realtà, alle applicazioni Compute-Intensive si sono man mano aggiunte quelle Data-Intensive che hanno un peso sempre maggiore e che richiedono tecnologie, capacità e competenze per gestire in maniera rapida ed efficiente l'enorme quantità di dati prodotta dai supercomputer, salvaguardando ciò che di più importante e vitale è per la ricerca: i risultati e le statistiche”*.

Un esempio sono le applicazioni Next Generation Sequencing nel settore della bioinformatica e biologia computazionale, che utilizzano sequenziatori in grado di realizzare un genoma umano in pochi giorni (le prossime generazioni di sequenziatori lo faranno in poche ore) arrivando a produrre decine di TB al giorno da elaborare, sintetizzare e archiviare in sicurezza, mettendo a dura prova l'infrastruttura di storage.

INFRASTRUTTURA STORAGE BASATA SU EMC

Al CASPUR, la capacità di storage sta aumentando sensibilmente ogni anno: i circa 15 TB complessivi disponibili nel 2008, nei tre anni successivi sono diventati più di 200 TB.

Per sostenere questo impegnativo trend di crescita senza penalizzare i livelli di servizio e supportando le prestazioni sempre maggiori dei supercomputer, il Consorzio ha deciso di sostituire l'infrastruttura storage esistente con una più avanzata che rispondesse a criteri di selezione basati su rigidi parametri di affidabilità, performance, disponibilità, flessibilità di configurazione, virtualizzazione del pool di storage e scalabilità elevata per aumentare i Terabyte disponibili senza interferire sull'operatività dei sistemi.

La gara, alla quale hanno partecipato i principali fornitori di tecnologie storage, è stata aggiudicata ad EMC che si è imposta per la migliore rispondenza ai criteri tecnici e per il rapporto costi/benefici.

I componenti dell'infrastruttura EMC, utilizzata sia dai supercomputer della struttura HPC sia dai server virtuali dell'ambiente VMware dedicati ai servizi per la Pubblica Amministrazione, sono:

- 3 sistemi storage **EMC CLARiiON Cx4-960** per una capacità complessiva di 1.2 PB espandibile fino a 6 PB. La possibilità di configurare i



sistemi con tipologie diverse di drive, permette di realizzare uno storage tiering per rispondere alle diverse esigenze di prestazioni delle applicazioni. L'80% dei dischi installati nei sistemi EMC CLARiiON del CASPUR sono SATA e il 20% Fibre Channel. A questi si aggiungono 5 TB di unità Flash con tempi di risposta costanti inferiori a 1 ms. Quest'ultima è una caratteristica esclusiva dei EMC CLARiiON e non disponibile su sistemi midrange di altri fornitori.

- 2 sistemi **EMC Connectrix** basati su tecnologia Cisco MDS 9605. Director multi protocollo che assicurano livelli di throughput da 8 Gb/s e scalabilità da 12 a 192 porte FC a 1, 2, 4 e 8 Gb/s.
- 2 sistemi **EMC VPLEX**, la soluzione di virtualizzazione dello storage disponibile sia in modalità local sia distributed federation per consentire, rispettivamente, la mobilità e l'accesso ai dati all'interno del data center, nonché tra data center diversi. Attualmente gli EMC VPLEX presenti nel CASPUR sono in local federation dando la possibilità di accedere ai dati su array anche eterogenei e realizzando una vera virtualizzazione dello storage. EMC VPLEX è stato utilizzato per migrare i dati dai precedenti sistemi storage ai nuovi EMC CLARiiON ridondandoli su più drive, funzionalità esclusiva non disponibile con soluzioni di altri fornitori.

“Dell'infrastruttura EMC abbiamo apprezzato e considerato vincenti alcune peculiarità tecniche e funzionali estremamente importanti per le attuali esigenze e per i nostri piani di crescita. La configurabilità e scalabilità del sistema EMC VPLEX ci permette di ampliare la capacità di storage in maniera trasparente, senza interferire su prestazioni e throughput complessivi, mettendoci a disposizione oltre 500 TB di spazio storage virtualizzato. Inoltre, la sua certificata integrazione con l'ambiente VMware è un valore determinante per i nostri piani futuri verso il Cloud Computing. Per quanto riguarda i director EMC Connectrix, basati su sistemi Cisco MDS, è indiscusso e riconosciuto nel mercato il loro grado di affidabilità. Così come altrettanto riconosciuta è la qualità e affidabilità degli storage EMC CLARiiON con i quali possiamo realizzare uno storage multi-tier, con prestazioni e costi di esercizio differenti, grazie alla configurabilità con drive SATA e FC e dischi flash allo stato solido”.

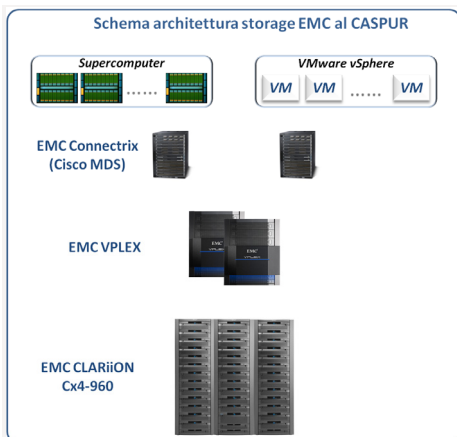
NICO SANNA
STRUTTURA HPC DI CASPUR

UN HPC ECOLOGICAMENTE SOSTENIBILE

Il sistema Jazz rappresenta la prima importante tappa del percorso che il CASPUR sta compiendo verso la realizzazione di una architettura Ict sostenibile, che permetterà nei prossimi 3 anni di abbattere del 50% il consumo energetico dei data center. È un risultato che il Consorzio vuole raggiungere con l'impiego di apparecchiature hardware sempre più 'verdi' e con la virtualizzazione, ottimizzando l'uso delle risorse e dei consumi.

Perfettamente in linea con questa strategia è l'implementazione dell'ambiente VMware che, nel caso del CASPUR, permette di consolidare fino a 10 macchine virtuali per ogni server fisico riducendo del 30-40% il consumo energetico complessivo.

Nell'ottica di green computing rientra anche l'infrastruttura storage EMC scelta da CASPUR. A parità di capacità, la tecnologia EMC CLARiiON Cx4 richiede un consumo energetico inferiore ai sistemi di generazione precedente. Da parte sua, EMC VPLEX, virtualizzando l'intero spazio storage disponibile permette di sfruttare meglio le risorse riducendo conseguentemente i consumi.



PREDISPOSIZIONE ALLA SOLUZIONE DI DISASTER RECOVERY

Il progetto, attualmente allo studio, di realizzare una soluzione di disaster recovery sarà facilitato dalla presenza di EMC VPLEX che in modalità distributed federation offre la mobilità dei dati, l'high availability e collaborazione ottimizzate tra due siti che si trovino a distanze sincrone (versione Metro) o asincrone (versione Geo).

IL PROGETTO CLOUD COMPUTING CON EMC E VMWARE

Nella veste di in-house provider e di partner tecnologico rispettivamente nei confronti delle università consorziate e della PA, il CASPUR sta definendo un catalogo di servizi in ottica Cloud Computing per rispondere alle esigenze sempre più pressanti di outsourcing e di delocalizzazione dei sistemi per contenere i costi di gestione dell'ICT.

“Per il Consorzio, il progetto di Cloud Computing è di fondamentale importanza strategica nell’area dei servizi ma, in prospettiva, anche per l’HPC” conferma Sanna, “Abbiamo individuato EMC e VMware come fornitori di riferimento, instaurando con loro uno stretto rapporto di partnership tecnologica e commerciale per individuare e definire le voci da inserire nel Catalogo dei Servizi Cloud e Infrastrutturali”.

Considerando la natura del mercato e la tipologia dei clienti cui è rivolta, l’offerta Cloud del CASPUR sarà declinata secondo le specifiche esigenze, con servizi che comprenderanno il noleggio di macchine virtuali anche in modalità pay per use, il noleggio di spazio disco garantendo la riservatezza dei dati e la sicurezza nell’accesso, la replica sincrona o asincrona dei dati via FC, l’archiviazione e il backup via IP, servizi RIS/PACS per il mercato ospedaliero. Altri servizi andranno ad aggiungersi nel tempo ampliando sempre di più l’offerta.

EMC², EMC, EMC CLARiiON, EMC VPLEX, EMC Connectrix sono marchi registrati di EMC Corporation. Tutti gli altri marchi o marchi registrati in questo documento sono proprietà delle rispettive aziende.