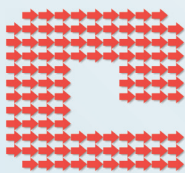


L'ACQUISIZIONE DI SISTEMI *HPC*: UN CASO ED ALCUNE NOTE GENERALI



di Carlo Maria Serio

Alla fine del 2008, il CASPUR ha rilasciato all'utenza un nuovo *cluster* per *HPC*. Il sistema presenta un ottimo livello di prestazioni ed un rapporto tra prestazioni e costo di particolare rilievo. Tale risultato è conseguenza di un accorto processo di dimensionamento ed acquisizione basato su una scelta tecnologica che ha guardato con particolare attenzione al quadro evolutivo dell'*HPC*.

Dott. Carlo Maria Serio
CASPUR
Responsabile del Settore HPC
c.serio@caspur.it

• Abstract

At the end of 2008, CASPUR has released to its users a new HPC system. The new system, built-up by ClusterVision, has 258 nodes with two quad-core AMD Opteron processors, for a total of 2.064 cores. Nodes are linked through a broad band and low latency Infiniband switch and have a total of 4.5Tbyte of dynamic memory. The system delivers 17.3 TFlops (Rpeak) and 13.6 TFlops (Rmax). Performances and total core-hours per year have grown 15 and 12 times respectively whilst cost/GFlops has declined to 1/40 of the previous one. According to CASPUR, the process adopted to acquire the system has been a best practice, with respect of technology and economics results, suitable for being used in future procurements. The article intends to illustrate such process, underlining main considerations and motivations in terms of technology, innovation, user needs adopted to define kind and size of the system as well as the parameters used to perform the choice. Some more is reported about world-wide HPC innovation scenario, stressing differences between HPC procurement in Europe and in the USA. Eventually, a couple of draft proposals to improve the effectiveness of overall HPC investments in Italy.

Nell'ultimo periodo del 2008 il CASPUR ha rilasciato all'utenza il nuovo sistema di *High Performance Computing (HPC)*, costituito da 258 nodi dotati ciascuno di due processori *AMD Opteron quad-core*, per un totale di 2064 *core*. I nodi sono connessi con tecnologia *Infiniband* ad alta banda e bassa latenza e sono dotati in totale di 4.5TByte di memoria dinamica.

La nuova *facility*, fornita da ClusterVision e battezzata *Matrix*, costituisce per il CASPUR un'importante novità sia per le prestazioni ed i costi, sia per il processo e le valutazioni che ne hanno governato la scelta. Guardando al grafico riportato in Figura 1 notiamo che, dal 1993 al 2006, il CASPUR ha acquisito un nuovo sistema principale ogni due-tre anni. Con un paio di eccezioni, ad ogni acquisizione la potenza del nuovo sistema è cresciuta di due o tre volte rispetto al precedente passando da 1 *GFlops* del 1993 ai circa 1,200 (1.2 *TFlops*) del 2006. L'aumento di

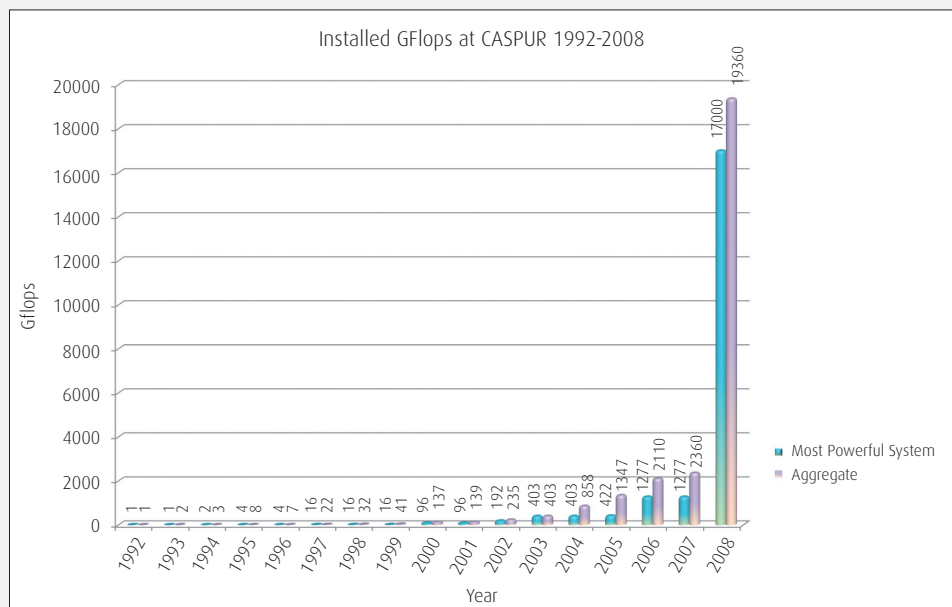
Figura 1

La *facility* HPC recentemente acquisita dal CASPUR ha una potenza quindici volte più elevata della precedente.

potenza e l'aggiornamento tecnologico derivante dai nuovi sistemi, hanno messo a disposizione degli utenti del CASPUR una piattaforma HPC sempre adeguata alle loro richieste. Il nuovo sistema, con 17.3 e 13.6 *TFlops* (*Rpeak* ed *Rmax* rispettivamente)¹ e 18 milioni di *ore/core* per anno, cresce rispetto al precedente di ben 15 volte in potenza e di 12 volte in capienza² di calcolo.

Ancora più significativo è l'andamento del costo/*GFlops*. Limitando l'analisi agli ultimi tre sistemi acquisiti e fatto uno il rapporto costo/*GFlops* del primo sistema (IBM Power 4, 2003), quello del secondo (IBM Power 5, 2006) scende ad un quarto mentre quello dell'ultimo vale solo un quarantesimo del secondo. Da notare che i valori non sono attualizzati.

Tali risultati sono solo in parte dovuti all'evoluzione dell'offerta, che presenta aumento delle prestazioni e diminuzione dei prezzi, ed infatti, una parte significativa è dovuta ai criteri adottati da CASPUR per la definizione della nuova *facility* e dalla procedura utilizzata per acquisirla. Ma procediamo con ordine e, prima di tutto, facciamo qualche considerazione su tecnologia ed innovazione nell'HPC.



L'HPC è un settore in cui la tecnologia deve esprimersi al meglio. A prescindere dalle varie architetture e stante l'impossibilità fisica e teorica di provvedere mediante un unico "iperprocessore" (in attesa dei *computer* quantistici), si risponde alle crescenti e puntuali necessità di calcolo mediante un insieme di processori interoperanti. Sia pure molto sommariamente, la velocità del singolo *computing element*, il loro numero, la memoria disponibile, la quantità di dati trasmissibili nell'unità di tempo (banda) ed il ritardo con cui si riesce ad accedere ad essi (latenza), costituiscono gli elementi che, più o meno integrati, determinano la potenza di calcolo complessiva di un sistema HPC.

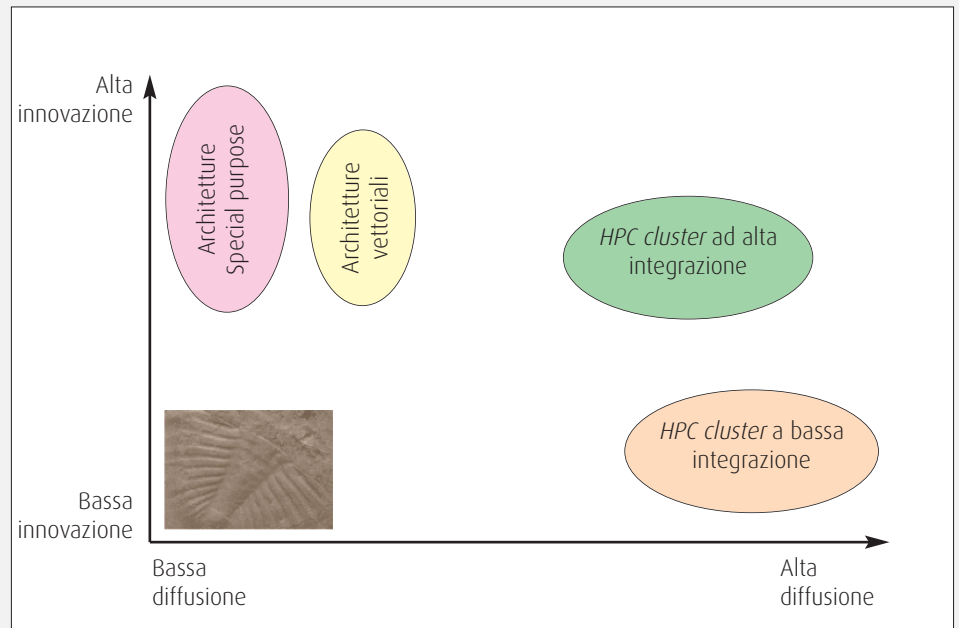
¹ Sono i valori di Theoretical Peak Performance e Max LINPACK Performance Achieved da Top500 list, Novembre 2008 (www.top500.org).

² Qui per "capienza" si intende il numero di *core* o *computing element* per 8.760, pari alle ore di un anno; è un'attribuzione arbitraria ma che consente una misura e richiama la traduzione di *capacity* fatta più avanti.

Come succede in altri campi dove tecnologia ed innovazione sono rilevanti, per l'*HPC* prevalgono due elementi: il costo delle soluzioni innovative ed il rischio di una loro rapida e prematura obsolescenza, entrambi decrescenti con la loro diffusione. Nella Figura 2 compaiono alcuni insiemi di soluzioni ed integrazioni *HPC* rappresentati in funzione della innovazione tecnologica e della diffusione. Le soluzioni altamente innovative hanno buona possibilità di sfociare in un vicolo cieco, scontando così un probabile destino di fossili tecnologici³, se non sono supportate da una significativa diffusione. Senza citare casi particolari, che ogni lettore documentato potrà richiamare facilmente alla mente, la storia dell'*HPC* è piena di soluzioni innovative cadute rapidamente nel dimenticatoio dei più ma che restano segnatamente presenti nella memoria di chi le ha incautamente adottate.

Figura 2

Soluzioni innovative che non conquistano larga diffusione presentano alti costi e rischi di rapida obsolescenza.

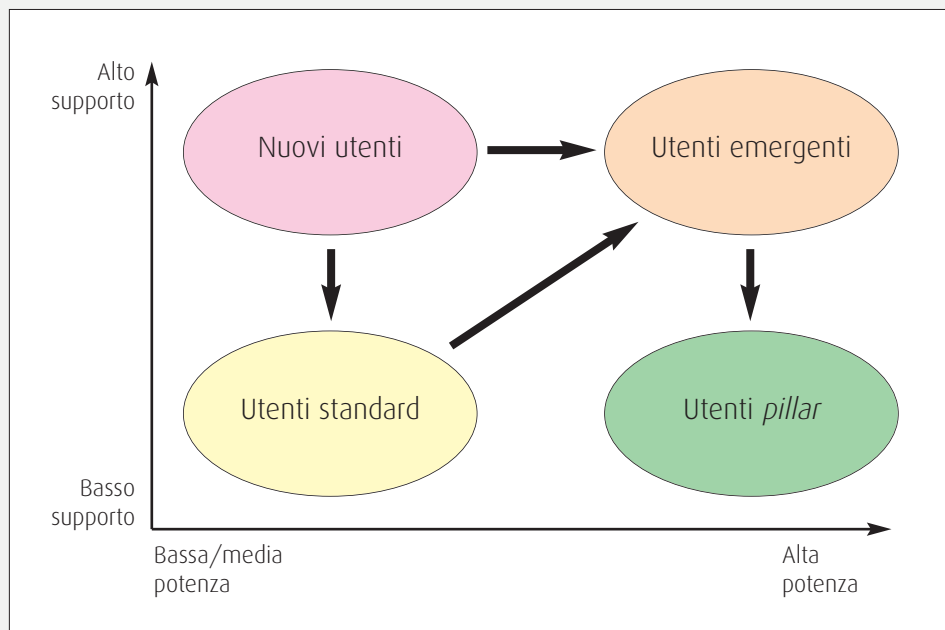


La rapida obsolescenza di una soluzione non sarebbe un gran problema se non presentasse lo svantaggio di un costo notevole, tanto più grande quanto maggiori sono gli investimenti richiesti. Ovviamente, quanto detto non ha pretesa di rigore, ma è solo una "testimonianza personale" basata sull'osservazione di fatti evidenti e, mi sembra, ricorrenti. Comunque va detto chiaramente che tutte le soluzioni molto innovative presentano rischi di prematura obsolescenza, ma anche che tale rischio è il prezzo da pagare per l'innovazione. Occorre perciò grande oculatezza, specialmente quando le risorse economiche sono limitate, di provenienza pubblica, ed il fine degli investimenti è la disponibilità all'utente di soluzioni valide e durature. Per quanto detto, e facendo ancora riferimento alla Figura 2, riteniamo che in Italia ed in Europa, gli investimenti *HPC* andrebbero effettuati principalmente negli insiemi rappresentati nella parte superiore destra dello schema, quella ad innovazione medio-alta ed ampia diffusione, riservando una quota so-

³ In realtà i "fossili" non rappresentano solo insuccessi nell'innovazione, sono anche soluzioni un tempo di relativo successo e diffusione, tenute in vita solo perché la loro sostituzione avrebbe costi o impatti non accettabili (*legacy*).

Figura 2

Tra gli insiemi di utenti HPC rappresentati in funzione della potenza di calcolo e del livello di supporto richiesto, solo quelli *pillar* necessitano di alta *capability*.



Gli utenti *pillar* sono quelli che spesso sviluppano autonomamente le loro applicazioni, definiscono l'ambiente operativo più opportuno e, tipicamente, hanno bisogno di prestazioni di punta per lunghi periodi. A volte, nel tentativo di spostare in avanti gli orizzonti della ricerca computazionale, si rivolgono alle tecnologie più innovative e, quando possono, contribuiscono a definirne lo sviluppo. L'insieme *pillar* è il solo che, a nostro avviso, può richiedere una *capability* molto elevata. Noi riteniamo che i centri HPC per gli utenti *pillar* debbano essere centri specifici, preferibilmente monotematici, ed in numero limitato, uno o due per l'Italia. Gli altri utenti, nel loro insieme, hanno bisogno di prestazioni più limitate, non esasperate e, specialmente gli utenti nuovi ed emergenti, necessitano anche di competenze a supporto dei loro progetti. In altre parole, gli utenti *pillar* hanno bisogno di *capability* HPC, dove la traduzione italiana migliore è "capacità" intesa come capienza e abilità insieme, tali da dare risposta alla particolare e significativa necessità di calcolo. Le simulazioni sul clima terrestre e sull'evoluzione dell'universo sono due esempi di problemi che, se affrontati su larga scala, richiedono *capability* HPC. Tutto il resto si può affrontare con adeguata *capacity*, tradotta come "capienza", ovvero volume adeguato a contenere più cose. La *capacity*, con una misura ridotta di *capability*, è fattore abilitante per i nuovi utenti, per quelli *standard* e per quelli *emergenti*. Tutti i centri HPC non specialistici, e



comunque non dotati di sistemi al vertice della *capability*, possono contribuire alla *capacity* complessiva dell'ecosistema *HPC*. Il loro numero e distribuzione sul territorio costituiscono un fattore fortemente abilitante all'aumento delle discipline e delle attività di ricerca suscettibili di avvalersi dell'*HPC*, anche e principalmente per la loro capacità di attrarre e supportare localmente gli utenti. Il modello che auspichiamo è quello di *capability* concentrata in pochi centri e di *capacity* distribuita sul territorio in un rapporto paritario in termini di potenza complessiva, meglio se il tutto è organizzato in maglie a grande accessibilità, secondo un modello di *grid*, per intenderci. In questo quadro, la scelta operata dal CASPUR nel dotarsi di un nuovo sistema, è stata quella di aumentare il proprio contributo prestazionale in un ambito di *capacity* e di mantenimento della necessaria copertura territoriale con adeguate competenze. Ultima considerazione, che introduciamo senza svilupparla, riguarda il maggiore ritorno sugli investimenti relativo al modello proposto, derivante dall'aumento dei ricercatori che fanno uso di *HPC* e, banalmente, da un reale risparmio economico.

Infine, illustro brevemente la procedura di acquisto adottata. Non uso il termine *procurement* perchè, a mio avviso, tale termine ha scarso significato in Europa dove, a parte opinabili eccezioni, non esistono costruttori di sistemi *HPC* e, considerando gli integratori, la tecnologia che utilizzano proviene quasi esclusivamente da oltreoceano. Per noi è impossibile fare *procurement* come lo fanno, per esempio, le varie strutture del Department of Energy o del Department of Defense degli Stati Uniti, le quali forniscono ai costruttori specifiche tali da poter essere soddisfatte solo con nuove soluzioni. In Europa si ha accesso solo a prodotti commerciali, definiti e realizzati altrove. Non possiamo andare in sartoria per decidere stoffe, taglio e rifiniture ma dobbiamo accontentarci di andare a scegliere tra i modelli e le taglie disponibili presso questo o quel negozio: manca l'*haute couture* e, dovendoci accontentare del *prêt-à-porter*, almeno facciamo in modo di ottenere il massimo ritorno economico.

Il CASPUR ha impostato la procedura di acquisto su tre elementi, tutti riconducibili all'obiettivo primario di stimolare una concorrenza vera e motivata e, quindi di massimizzare il risultato.

Il primo elemento ha riguardato l'impostazione della procedura, avendo l'intento d'interessare molti possibili fornitori. In teoria, per ottenere un buon risultato al riguardo è sufficiente dichiarare che tutte le possibili coppie di fornitore-soluzione che rispettano i requisiti richiesti, avranno pari opportunità. Nel concreto, bisogna indicare requisiti che non escludano a priori nessun potenziale fornitore o, peggio, non conducano a questo o a quello. In questa ottica, ci si è sforzati di stilare un capitolato rigoroso, semplice e chiaro e di gestire tutte le relative comunicazioni in modo tempestivo, corretto ed esauriente, anche riguardo le modalità connesse alla valutazione.

Il secondo elemento è coinciso con lo standard della soluzione tecnica richiesta, definito, riferendoci a quanto detto prima, in un ambito di tecnologia diffusa e privo di dettagli discriminanti. Per garantire la massima apertura, era consentito al concorrente di proporre anche due o più soluzioni basate su differenti topologie di *networking*, numero o tipo di processori purchè fossero salvi i requisiti comuni.

Il terzo ed ultimo elemento ha riguardato il miglioramento economico. Stabilito il *budget* e la configurazione di massima, invece di un ribasso sul prezzo offerto, è stato chiesto ai partecipanti di proporre configurazioni migliorative a prezzo inalterato. Abbiamo infatti ritenuto che, per le aziende, sia più facile lavorare sul solo calcolo dei margini se il presunto ricavo risulta predefinito: chi conosce le logiche di *quotation* aziendali può comprendere questo approccio. Ad ogni modo, riteniamo di non aver scoperto nulla di nuovo ma, anche, di aver lavorato affinché quanto già noto contribuisse realmente a produrre il miglior risultato possibile.

In sintesi, riteniamo che l'acquisizione del nuovo sistema HPC del CASPUR sia stata una *best practice* in termini di gestione, scelta tecnologica e dimensionamento. Ne diamo pubblicità nella presunzione che possa risultare utile anche ad altri. Vorremmo infine concludere con una doppia proposta. La prima è quella di condividere le *best practice* (e magari anche le *worst ones*) tra tutti gli interessati, superando un consolidato atteggiamento di riservatezza dovuto forse a qualche ragione non più giustificabile. La seconda è quella di arrivare, facendo davvero sistema, ad una rotazione delle acquisizioni dei sistemi nei centri HPC che preveda, per esempio, anche la valutazione comune di soluzioni particolarmente innovative. A mio avviso ci sarebbero enormi vantaggi per tutti, *in primis* per la comunità scientifica. Non è il caso di elencarli ma, se appena si riescono ad intuire, è bene iniziare a discuterne, magari non oggi, ma domani sì: sarebbe davvero importante!