

# COSMOLAB

*Consorzio per il supercalcolo, la modellistica  
computazionale e la gestione di grandi database*

Al Presidente del consorzio Cosmolab  
Prof. Giuseppe Mazzarella

e, p.c.,

Al Presidente del CS Cosmolab  
Prof. Gianni Mula

Al Responsabile Scientifico del Progetto CyberSAR  
Dr. Albero Masoni

Ai Referenti di INAF, CRS4, UniCA Ingegneria, UniSS  
Ignazio Porceddu  
Lidia Leoni  
Gianni Fenu  
Marco Masia

Cagliari, 10 novembre 2008

Oggetto: criticità del cluster UC1 – hardware IBM

Con la presente desidero segnalarti ufficialmente un problema critico a cui è soggetto il cluster UC1 – infrastruttura di calcolo del polo CyberSAR di Monserrato – derivante dalla inaffidabilità hardware dei server IBM System x3455 type 7986 in nostro possesso. Tale inaffidabilità è stata riscontrata sin dall'inizio della messa in funzione dei server e continua a manifestarsi come meglio specificato nel seguito.

La sola statistica delle richieste di assistenza hardware è sufficiente a sottolineare come il problema vada ben oltre la fisiologica occorrenza di guasti a cui sono normalmente soggetti i dispositivi elettronici; infatti, sui 100 server IBM System x3455 type 7986 del polo CyberSAR di Monserrato

- sono state effettuate complessivamente circa 50 richieste di assistenza per problemi hardware, risolte per la maggior parte con la sostituzione della componente guasta: l'occorrenza dei guasti che hanno coinvolto dischi, alimentatori e schede di rete rientra abbondantemente nei limiti fisiologici, l'occorrenza dei guasti relativi alla memoria RAM è stata superiore ma tutto sommato tollerabile, l'occorrenza dei guasti che hanno coinvolto le planar è inaccettabile, come si evince dai numeri indicati nel punto successivo
- circa 30 richieste su 50 hanno comportato la sostituzione della planar (nome politicamente corretto con cui IBM identifica la scheda madre dei server)

I problemi più frequentemente riscontrati in relazione alle planar difettose sono

1. riavvio improvviso del server, a volte continuo
2. spegnimento improvviso del server
3. mancato avvio del server

Nonostante l'esecuzione di test diagnostici effettuata su indicazione di IBM (l'ultima risale a pochi giorni fa), non ci è stata fornita da IBM alcuna causa precisa alla base dei problemi evidenziati dalle

# COSMOLAB

*Consorzio per il supercalcolo, la modellistica  
computazionale e la gestione di grandi database*

planar. In effetti, i malfunzionamenti delle planar si manifestano all'improvviso con i sintomi sopra descritti, ma non lasciano traccia nei registri degli eventi a livello di sistema operativo o a livello hardware. Inoltre, tali malfunzionamenti si sono presentati su alcuni server sin dall'inizio, mentre su altri si sono manifestati in tempi successivi, e tutt'ora continuano a contagiare nuovi server in maniera imprevedibile (tranne per il fatto che l'occorrenza è decisamente più bassa sui server a cui è stata sostituita la planar, ma non nulla), causando problemi che vanno ben oltre il blocco del singolo server, come descritto più avanti. A ciò si aggiunge il fatto che la maggioranza delle parti di ricambio inviate da IBM per la sostituzione non è nuova (per le planar la quasi totalità), si tratta di parti usate come si evince facilmente dal confezionamento, dall'etichettatura delle scatole, dal contenuto del registro degli eventi hardware della planar; un'ulteriore conferma è il fatto che il numero di serie di una planar inviata per la sostituzione presso il cluster UC1 è risultato lo stesso del numero di serie di una planar sostituita (perchè presumibilmente guasta) presso il cluster OA1 – infrastruttura di calcolo del polo CyberSAR INAF. In più di un caso il tecnico IBM intervenuto per la sostituzione di una planar ha dovuto fare una doppia sostituzione: la prima planar sostituita è risultata anch'essa non funzionante alla riaccensione del server e dunque è stato necessario procedere con una seconda sostituzione. Tali malfunzionamenti non possono essere imputabili all'installazione dei server in un ambiente non adeguato; infatti, nella sala CED del polo CyberSAR di Monserrato,

- l'alimentazione elettrica dei server è erogata da UPS adeguatamente dimensionati e ridondati
- la temperatura dell'aria in ingresso ai server è costante e pari a circa 19-21 gradi centigradi
- l'umidità relativa dell'aria in ingresso ai server è costante e pari a circa il 50-55%

Inoltre, il firmware dei server è stato aggiornato all'ultima versione rilasciata da IBM alla data della messa in produzione dei server e il sistema operativo installato sui server è quello indicato nel capitolato di gara sottoscritto dal fornitore.

I malfunzionamenti descritti in precedenza hanno finora avuto, nella maggior parte dei casi, un effetto limitato alla non operatività del solo server oggetto del guasto (fortunatamente i server System x3455 vengono utilizzati solo come nodi di calcolo): questo ha comportato l'interruzione dell'eventuale job in esecuzione sul server; è bene comunque osservare che per un job parallelo su N server basta il guasto di uno qualunque degli N server perchè l'intero job venga interrotto.

Purtroppo, però, il malfunzionamento di uno dei server avvenuto in data 9/11/2008, manifestatosi con ripetuti riavvii improvvisi senza uno shutdown ordinato del sistema operativo, ha evidenziato una criticità che ha portato all'inoperatività dell'intero cluster con l'interruzione e la conseguente perdita di tutti i job in esecuzione. Il successivo ripristino dell'operatività del cluster ha richiesto l'intervento manuale di un operatore, e richiederà una nuova esecuzione da parte degli utenti dei job interrotti. Poichè questa è la seconda volta in meno di tre mesi, non si può più catalogare come un fatto isolato e dunque trascurabile, ma diventa un problema di affidabilità e di stabilità dell'intera infrastruttura di calcolo. Un'analisi approfondita ha rivelato che il problema è legato al software IBM GPFS v3.2 che gestisce il filesystem del cluster e permette la condivisione del sistema di storage centralizzato IBM DS4800 a tutti i server del cluster. Tale software adotta una modalità di funzionamento distribuito, e in caso di connettività non stabile con i server del cluster – un server che continua a riavviarsi non

# COSMOLAB

*Consorzio per il supercalcolo, la modellistica  
computazionale e la gestione di grandi database*

garantisce certo una connettività stabile – può perdere la sincronia. In quest'ultimo caso, il filesystem viene smontato da tutti i server del cluster e rimontato dopo il tempo necessario ad effettuare il recovery. Quando il filesystem viene smontato, i server non hanno più accesso allo storage condiviso e l'esecuzione delle applicazioni si interrompe.

Il supporto tecnico IBM è sostanzialmente a conoscenza dei malfunzionamenti delle planar descritti all'inizio, a causa delle innumerevoli interazioni che abbiamo con esso, ma non conosce le implicazioni che questi possono avere per il Consorzio, e d'altro canto il suo mandato non va oltre la sostituzione delle planar guaste, caso per caso e in base alle nostre esplicite richieste di assistenza.

Ti sollecito pertanto ad informare prontamente anche i nostri referenti commerciali IBM in modo da arrivare al più presto ad una soluzione definitiva prima della scadenza della garanzia, per esempio attraverso la sostituzione con planar nuove di tutte le planar non ancora sostituite.

Se questo stato di cose dovesse permanere, è chiaro che il Consorzio non è in grado di garantire l'affidabilità e la stabilità dell'infrastruttura di calcolo del polo CyberSAR di Monserrato e questo comporta, fra l'altro, l'ovvia impossibilità di poter fornire in futuro un servizio di calcolo a pagamento. Inoltre, se IBM non trova una adeguata soluzione al problema, ti annuncio fin d'ora la mia completa opposizione all'acquisizione di qualsiasi altro sistema di calcolo IBM all'interno del progetto CyberSAR o in qualsiasi altro progetto in cui dovessi essere coinvolto.

Un caro saluto,

Il Responsabile di OR1-A1.1  
(Andrea Bosin)