

**Relazione sulle esigenze di hardware e di personale necessarie alla  
realizzazione di una infrastruttura di analisi locale (Tier3) per  
l'esperimento Atlas a Genova**

18 giugno 2009

Alessandro Brunengo

## Analisi dei requisiti hardware e software

In base alla discussione avuta all'ultima riunione del gruppo Atlas di Genova, l'installazione di una infrastruttura di analisi locale (Tier3) per l'esperimento Atlas ha alcuni requisiti che devono essere soddisfatti:

- **infrastruttura Grid:** e' quasi implicito che deve essere disponibile localmente una infrastruttura Grid connessa alla Grid di produzione INFN, con tutti i suoi componenti essenziali (CE, SRM, UI), e deve essere mantenuta operativa.
- **storage:** una opportuna accessibilita' delle aree dati dell'esperimento richiede che tali aree debbano essere accessibili sia tramite interfaccia SRM (e quindi integrate nella infrastruttura locale Grid) che direttamente dai nodi di analisi, che allo stato attuale non e' ancora definito se debbano essere integrati nella infrastruttura Grid o meno, ma le cui risorse devono essere necessariamente accessibili anche via batch system locale, o eventualmente in modalita' interattiva. La soluzione attualmente piu' efficiente per realizzare questo e' l'utilizzo di un SRM basato su StoRM, che si appoggi su un file system di rete accessibile direttamente anche dai nodi del Tier3. Date le caratteristiche di StoRM, tale file system deve essere uno dei due file system paralleli attualmente supportati: GPFS o Lustre. In questo modo i dati potranno essere trasferiti via comandi SRM dall'ambiente Grid remoto, per poi poter essere analizzati dai job tramite protocollo *file*.
- **rete locale:** il dimensionamento della potenza di calcolo del Tier3 non e' ancora definito nel dettaglio, ma ragionevolmente sara' costituito da almeno una decina di nodi di calcolo su cui dovranno girare contemporaneamente un numero di job pari al numero di core disponibili. La tecnologia attuale offre oggi 8 core per nodo di calcolo, ed in futuro potranno aumentare. Poiche' le esigenze di banda per ogni job e' valutata intorno ai 20 MB/s, ogni singolo nodo oggi dovrebbe disporre di 1 Gbps di banda verso lo storage, con un aggregato che potrebbe richiedere oltre 10 Gbps. L'infrastruttura di rete deve pertanto poter offrire un backbone ed un aggregato verso i dischi in tecnologia 10 GE.
- **rete geografica:** la banda necessaria verso Garr non puo' essere attualmente quantificata, ma e' ragionevole prevedere esigenze che comporterebbero un incremento della banda attuale (BGA: 20 Mbps, BEA: 50 Mbps) anche di un ordine di grandezza.

## Analisi della attuale infrastruttura

Al fine di valutare la sostenibilita' ed i costi necessari per la realizzazione del Tier3, segue una analisi della configurazione attuale delle infrastrutture di calcolo della Sezione di Genova, in relazione agli aspetti citati nel precedente paragrafo.

- **infrastruttura Grid:** il Servizio Calcolo della Sezione ha installato e mantiene una infrastruttura di calcolo Grid connessa alla Grid di produzione INFN, attualmente dotata di 22 core (presto 48) e 4 TB di spazio disco gestiti tramite SRM StoRM con backend GPFS.
- **storage:** il Servizio Calcolo ha una infrastruttura GPFS in produzione da quasi tre anni che ospita la maggior parte dei volumi locali non solo Grid (~80 TB complessivi) in configurazione totalmente ridondante, acceduta via Storage Area Network su protocollo Fibre Channel, e dispone del necessario know-how per il suo mantenimento ed il suo sviluppo, sia per l'hardware (SAN) che per il software (GPFS).

Disponendo di tale configurazione, al momento di migrare lo storage element dal classic SE ad SRM e' stata scelta la soluzione StoRM su GPFS, che e' attualmente in produzione. I volumi Grid ed i volumi centrali – tra i quali anche l'attuale spazio disco di Atlas – fanno parte di cluster GPFS differenti, ma l'integrazione dei due volumi non e' tecnicamente un problema insormontabile. Per quanto riguarda le prestazioni verso lo storage, il sistema disco attuale e' stato testato fornire prestazioni oltre i 7 Gbps con accesso prevalentemente sequenziale. E' pero' probabile che in sede di incremento di spazio disco per le esigenze del Tier3 dovra' essere acquistato hardware opportuno a garantire le prestazioni necessarie, e questo dovra' essere fatto in comune accordo tra l'esperimento Atlas ed il Servizio Calcolo.

- **rete locale:** l'attuale backbone della rete locale e' in tecnologia GE, ma e' stato gia' finanziato da CCR l'acquisto di una scheda con interfacce 10 GE da inserire nel centro stella per l'upgrade a questa tecnologia. L'acquisto verra' fatto entro la fine dell'estate. Manca invece la disponibilita' di uno o due disk server con interfaccia 10 GE da connettere al centro stella per portare i 10 Gbps fino alla Storage Area Network, e disporre di server di disco dedicati per le esigenze del Tier3. Andra' anche programmato l'acquisto di una nuova scheda a 48 porte GE da installare sul centro stella per la connessione dei WN (che potranno eventualmente essere connessi a 2\*1 Gbps configurando per ciascuno due interfacce di rete in bonding) o in alternativa, uno switch a 48 porte 100/1000 in rame con uplink a 10 GE.
- **rete geografica:** la connettivita' tra la Sezione di Genova ed il PoP Garr di Genova e' attualmente realizzata su fibra ottica di proprieta' dell'Ateneo Genovese, concessa in utilizzo gratuito. Il protocollo attualmente utilizzato e' FastEthernet (100 Mbps), su cui e' ritagliata una banda garantita di 20 Mbps ed una BEA di 50 Mbps, ma comunque capace di trasferire dati alla sua capacita' massima di 100 Mbps. Un eventuale upgrade della linea a 1 Gbps e' realizzabile dal lato nostro senza dover cambiare apparato di routing, mentre le modifiche necessarie lato Garr non sono note. L'eventuale upgrade andra' ovviamente concordato con Garr in base alle reali esigenze locali, ma non si vedono impedimenti tecnici insormontabili.

In base a quanto scritto l'infrastruttura di rete locale a fine estate, dopo l'acquisto della scheda con porta a 10 GE per il centro stella, sara' idonea a supportare il Tier3. Per quanto riguarda la configurazione software, la piattaforma in produzione e' quella idonea a realizzare il Tier3. Deve essere fatta l'integrazione dei volumi Grid con i volumi locali di Atlas, cosa che comportera' un periodo di lavoro quantificabile in tre settimane (essenzialmente per via della ricollocazione dei volumi di Atlas e di riconfigurazione degli account usati dai WN in Grid), che sara' fatta una volta che il gruppo avra' definito se procedere con la configurazione del Tier3.

Per la rete geografica, si rimanda ad una analisi piu' puntuale da fare in collaborazione con Garr quando l'utilizzo della linea attuale si sara' dimostrato insufficiente.

## Hardware da acquistare e valutazione economica

In base a quanto visto nei paragrafi precedenti, gli acquisti hardware da effettuare per gli apparati di l'infrastruttura dedicati al Tier3 sono i seguenti:

- due server di disco dotati di scheda Fibre Channel dual head e di scheda di rete 10 GE. Una configurazione attuale (biprocessore quad core con 16 GB di RAM) ha un costo di circa 7 K€ per server.
- una scheda a 48 porte 100/1000 in rame per il centro stella, che ha un costo valutabile intorno ad 8/10 K€ per la connettivita' dei WN. In alternativa uno switch 48 porte 100/1000 con uplink a 10 GE, di costo paragonabile.

A questo vanno aggiunti naturalmente i WN necessari, e l'eventuale storage aggiuntivo rispetto a quello oggi disponibile. Non saranno necessari ulteriori upgrade di hardware infrastrutturale per ospitare disco e nodi di calcolo.

Naturalmente tutto quanto detto ha significato nella ipotesi che le infrastrutture di alimentazione e di raffreddamento della sala macchine della Sezione possano essere interamente rifatte, come da progetto presentato alla Giunta Esecutiva al fine di ottenere un finanziamento dedicato allo scopo. Data la attuale situazione degli impianti, ed in assenza di tali lavori di ristrutturazione, infatti, non sarà possibile garantire nemmeno la possibilità di installare oggetti aggiuntivi a quelli oggi presenti.

## **Esigenze di personale per il setup ed il mantenimento della infrastruttura**

L'infrastruttura Grid richiede per la sua manutenzione, il 50% di un tecnologo. Oggi questo compito è coperto da un tecnologo precario del Servizio Calcolo (Corosu) che, pur essendo stabilizzabile, non è stato stabilizzato e non lo sarà entro la scadenza delle stabilizzazioni, a fine 2009.

Benche' ci siano motivi per pensare che la sua posizione sarà resa stabile nel 2010, ciò è oggi solo una ipotesi.

L'eventuale indisponibilità di questa persona renderà il Servizio Calcolo inabile a garantire il funzionamento delle infrastrutture di calcolo scientifico, compresa quella Grid. Un eventuale reintegro con altra persona di pari capacità ed attitudini richiederà un periodo di apprendistato che valuto essere almeno sei mesi per essere operativo.

Queste sono considerazioni che vanno tenute in conto nella pianificazione di un Tier3 in Sezione.

Detto questo, la realizzazione di un Tier3 per l'esperimento Atlas richiede assolutamente una collaborazione da parte di un tecnologo o ricercatore di esperimento, che possa collaborare con il Servizio Calcolo per il setup e la analisi di potenziali problemi dal lato utente (quindi mantenimento delle installazioni di software, operazioni di test di funzionalità e di test di prestazioni, collaborazione per attività di problem tracking con il Servizio Calcolo). Il carico di lavoro necessario è difficile da valutare, ma si può dare indicativamente un valore pari al 30% di una persona, con un impegno maggiore nella fase iniziale.