Capitolo 9

Conclusioni

In questa tesi è stata presentata l'analisi e la progettazione di un sistema, chiamato DEMI, per la programmazione di job, costituiti da unità computazionali elementari opportunamente combinate tra loro. Il sistema gestisce poi l'esecuzione di questi job in un ambiente di calcolo distribuito ed eterogeneo.

Il sistema è stato progettato soprattutto per poter gestire le elaborazioni tipiche dell'ambiente della fisica delle alte energie (senza riferimento ad uno specifico esperimento), ma esso è comunque facilmente "esportabile" anche ad altri domini.

È stato poi realizzato un primo prototipo del sistema, che già implementa la maggior parte dei requisiti definiti inizialmente. Il sistema è infatti in grado di gestire in maniera efficiente elaborazioni, anche complesse, in un ambiente costituito da workstation Unix e/o PC con sistema operativo Windows NT. Il sistema utilizza i diversi calcolatori in maniera uniforme e del tutto trasparente per gli utenti, che considerano DEMI come un unico, potente sistema di calcolo.

È importante sottolineare come, a differenza di sistemi già esistenti, DEMI è flessibile, scalabile, e facilmente configurabile: il sistema infatti non è specifico ad un particolare ambiente di calcolo, o ad un particolare tipo di elaborazione, e può essere adattato per le specifiche esigenze che si devono considerare.

I vantaggi più evidenti sono comunque per l'utente finale, che deve solo preoccuparsi di definire l'elaborazione da eseguire, e attendere che gli vengano forniti i relativi risultati: non deve preoccuparsi di gestire i vari dettagli implementativi. Altro elemento di novità è che anche l'attività di gestione da parte dell'amministratore è minima: le varie funzionalità sono del tutto "automatizzate". Il sistema è anche self managing, in quanto è in grado di gestire autonomamente eventuali failure, senza l'intervento dell'amministratore.

Con riferimento all'ambiente della fisica delle alte energie, le elaborazioni che si devono considerare richiedono spesso molte risorse (specialmente cpu e memoria), e ci sono quindi problemi ad usare le workstation utilizzate dagli utenti per questi processing. Attualmente ci sono due possibili approcci per l'esecuzione di questi job: acquistare workstation da dedicare solo a queste attività, o usare le workstation degli utenti solo in determinati periodi, quando si prevede che non ci sia attività di tipo interattivo (sostanzialmente di notte e nei fine settimana). In entrambi i casi ci sono problemi da considerare: il costo da sostenere per l'acquisto di nuovi calcolatori, il fatto che comunque di notte e nei weekend ci possono essere elaborazioni degli utenti, l'utilizzo non ottimale dei vari computer, le cui potenzialità non vengono del tutto sfruttate.

Con il sistema DEMI, grazie alla sua integrazione con un sistema di load sharing, questi problemi vengono superati: si garantisce il massimo utilizzo dei vari calcolatori, senza comunque interferire con le normali attività degli utenti.

9.1 Sviluppi futuri

Il sistema realizzato può essere completato e integrato sotto diversi aspetti.

Potranno essere fatte, innanzittutto, modifiche e raffinamenti per migliorare l'affidabilità e l'efficienza complessiva. Ad esempio, come spiegato nel capitolo precedente, può essere modificato l'algoritmo di scheduling locale, in modo da ottimizzare le prestazioni del sistema.

È poi necessario implementare un'adeguata interfaccia grafica, in modo da rendere semplice l'utilizzo del sistema da parte degli utenti e degli amministratori. Particolarmente curata dovrà essere l'implementazione dell'interfaccia grafica per la definizione di task e di produzioni. Potrà essere usato il formalismo proposto in questa tesi, rappresentando quindi i task attraverso ellissi, i data set attraverso rettangoli, e le produzioni attraverso grafi, in cui i vari task vengono combinati tra loro.

Altro possibile lavoro futuro è il porting dei vari moduli del sistema (Global Interface, Local Interface, DEMI Global Control System, DEMI Local Control System) alla piattaforma Windows NT: questa non dovrebbe essere, comunque, un'attività molto complessa, in quanto non si sono utilizzati particolari feature implementate solamente per le architetture Unix.

Si può poi prevedere l'integrazione del sistema DEMI con altri sistemi di load sharing, oltre a LSF, che abbiano comunque tutte le funzionalità richieste per la gestione di batch job.

È poi necessario l'impiego di un DBMS più efficiente e affidabile rispetto al sistema MySQL utilizzato nel prototipo realizzato. Questo DBMS deve anche essere in grado di gestire grossi volumi di informazioni. In praticamente tutti gli esperimenti di fisica delle alte energie si è scelto di adottare, per la memorizzazione dei vari dati, *Objectivity/DB*, un DBMS a oggetti, implementato su molte piattaforme (la maggior parte dei sistemi Unix, e Windows NT), e che permette di definire database distribuiti tra più calcolatori, posti anche in siti diversi [42]. Attualmente è in fase di definizione un'interfaccia tra Objectivity/DB e *HPSS* (*High Performance Storage System*) [58], un sistema software in grado di gestire in maniera gerarchica un volume di dati molto elevato (dell'ordine dei Pbyte), con ottime prestazioni [56]. Si può quindi valutare se Objectivity/DB può essere un utile strumento da impiegare nel sistema DEMI.

Per completare la realizzazione del sistema sarà infine necessario progettare e implementare il Global Scheduler, in modo da poter gestire le elaborazioni in un ambiente di calcolo distribuito geograficamente tra più siti.

Saranno poi gli utenti finali a suggerire migliorie o nuove funzionalità da implementare.

Ringraziamenti

Sono molte le persone che voglio ringraziare, in quanto, in forme e modi diversi, hanno dato il loro contributo alla realizzazione di questo lavoro.

Innanzittutto voglio ringraziare il relatore, dott. Salvatore Orlando, e i correlatori, il dott. Mirco Mazzucato e il prof. Roberto Zannoni, sempre disponibili a mettere a disposizione le loro competenze, e che mi hanno dato numerosi consigli e suggerimenti, di cui spero aver fatto buon uso.

Desidero ringraziare anche il dott. Jean Claude Marin e il dott. Franco Carena del Cern di Ginevra: dalle discussioni che ho avuto con loro ho sempre imparato qualcosa di nuovo.

Un caloroso ringraziamento anche a tutti i docenti del corso di laurea in scienze dell'informazione: in questi anni sono sempre stati tutti estremamente disponibili nel dedicarmi parte del loro tempo, andando ben oltre i loro obblighi.

Un ringraziamento speciale per i miei colleghi e amici del servizio calcolo dell'INFN di Padova: Federico Balbi, Roberto Ferrari, Massimo Gravino, Fabio Montecassiano, Ivo Saccarola, Lorenzo Turicchia, e in particolare Fulvia Costa: lavorare con loro è sempre stata ed è tuttora un'esperienza estremamente positiva. Sono grato anche a quelli che sono stati i responsabili del servizio calcolo in questi anni: il prof. Maurizio Morando, il dott. Roberto Stroili, il dott. Mauro Morandin e il dott. Michele Michelotto, che, nonostante i miei impegni di lavoro, mi hanno sempre sostenuto nel proseguimento degli studi.

Ringrazio poi tutti gli altri amici dell'INFN di Padova, in particolare Gabriele Rampazzo e Mario Zago che, oltre ad offrire la loro simpatia, hanno contribuito non poco alla realizzazione di questo lavoro.

Un grazie anche a tutti i membri dei gruppi DELPHI e CMS di Padova, in particolare il prof. Ugo Gasparini, il dott. Ivano Lippi, la dott.ssa Marina Passaseo e il dott. Enrico Piotto.

Infine, ma non per ultimo, ringrazio la mia famiglia, che non mi ha fatto mai mancare il suo sostegno.