



CONSIGLIO NAZIONALE DELLE RICERCHE

**PROGRAMMA SHORT - TERM MOBILITY
Anno 2008**

**Relazione sull'attività di ricerca svolta
presso il CERN di Ginevra
dal 19 Ottobre 2008 al 8 Novembre 2008.**

Luca Frosini

Istituto di Scienza e Tecnologie della Informazione "Alessandro Faedo"

Consiglio Nazionale delle Ricerche

Area della Ricerca CNR di Pisa

Via G. Moruzzi, 1 – 56124 Pisa, Italy

+39 050 3153473

luca.frosini@isti.cnr.it

SCOPO DEL PROGRAMMA

L'attività di ricerca svolta presso il CERN di Ginevra prevedeva il consolidamento e il potenziamento del D4Science Monitoring Tool al fine di migliorare il supporto che questo strumento fondamentale fornisce sia agli utenti che ai sistemi aventi il compito di gestire l'operatività di una e-Infrastructure.

Le e-Infrastructures sono sistemi complessi, aperti e distribuiti, consistenti di risorse eterogenee e dinamiche, il cui stato concorre alla definizione dello stato dell'intero sistema. Poter disporre di un quadro chiaro e sempre aggiornato dello stato di una e-Infrastructure costituisce un passo fondamentale per garantire la sua operatività e la qualità del servizio. Maggiore è il supporto per il monitoraggio, più "intelligente" e autonomo risulta il sistema dal punto di vista delle potenzialità operative.

D4Science è un progetto attualmente in corso nell'ambito del 7° Programma Quadro - Information Society Technologies dell'Unione Europea. Esso è focalizzato sullo sviluppo di una e-Infrastructure "production level" che fornisca alle comunità di "Environmental Monitoring and Fishery" e "Aquaculture Resources Management" un Virtual Research Environment. Uno degli obiettivi del progetto è quello di ridurre al minimo la necessità di intervento umano per gestire l'operatività del sistema.

Il Monitoring Tool di D4Science prima della mia visita al CERN era il risultato del consolidamento ed estensione di un servizio inizialmente nato per servire le necessità di un sistema di "test-bed". Il consolidamento e l'estensione ha avuto lo scopo di rendere il servizio capace di fornire le funzionalità attese nel contesto di una infrastruttura di produzione, cioè di migliorare l'affidabilità e la stabilità del servizio. Durante il periodo di visita l'attività è stata indirizzata principalmente a migliorare il servizio sia dal punto di vista della completezza funzionale che da quello dell'usabilità.

Per raggiungere tale obiettivo sono state effettuate le seguenti attività:

- Analisi dei requisiti operativi.
- Breve survey degli strumenti e tecnologie esistenti.
- Progettazione delle estensioni del Monitoring Tool.
- Implementazione prototipale delle estensioni previste.

Al momento della stesura della proposta, erano previsti almeno due requisiti operativi: il supporto di "multiple focused views" dell'infrastruttura e il supporto di un monitoring pro-attivo.

L'attività di ricerca svolta al CERN è iniziata con un giorno di lavoro a contatto con l' Ing. Pedro Andrade, per la pianificazione delle attività da svolgere durante le tre settimane del programma. L' Ing. Andrade è stato molto disponibile fin dall'inizio e ha cercato da subito di introdurmi nell'ambiente del CERN, assegnandomi una postazione in un ufficio il più vicino possibile al suo e operando attivamente per la gestione delle pratiche burocratiche a me relative.

1° settimana

Dopo il giorno di pianificazione delle attività, come concordato con l' Ing. Andrade ho iniziato lo studio dei requisiti operativi del tool di monitoring.

I requisiti fondamentali che sono stati individuati sono:

- Necessità di 3 differenti tipi di viste, una per ogni livello di astrazione infrastrutturale: GHN based, VRE based, VO based.
- Possibilità di gestire ogni vista in base alle caratteristiche peculiari della vista stessa.

Sono state quindi individuate le caratteristiche peculiari per la gestione della vista.

In particolare sono state identificate le seguenti peculiarità:

- Vista GHN based:
 - Numero di Running Instance attive sul GHN.
 - Numero di CPU a disposizione del GHN.
 - Spazio disco disponibile per il GHN.
 - Memoria Virtuale disponibile per il GHN.

Tali caratteristiche permettono di valutare la bontà di un GHN per poter essere selezionato per il deployment dinamico di una nuova Running Instance anche in base ai requisiti della Running Instance stessa.

- Vista VRE based:
 - Numero di collezioni di oggetti disponibili.
 - Numero di oggetti totali disponibili nel VRE.
 - Tipologie di indici disponibili per il VRE.

Tali caratteristiche permettono di rilevare lo stato di una VRE in termini di risorse fruibili dai membri del VRE stesso.

- Vista VO based:
 - Servizi Disponibili
 - Running Instance attive
 - Numero di GHN

Tali caratteristiche permettono di valutare lo stato della VO in termini di servizi disponibili e in effettiva esecuzione al fine di poter massimizzare l'utilizzo delle risorse stesse o incrementarne le capacità qualora risultassero insufficienti.

Una volta individuati i requisiti operativi è stato studiato lo stato dell'arte e sono state individuate tra le tecnologie esistenti quelle utilizzate nell'ambito del progetto EGEE per il monitoraggio dell'infrastruttura grid. Tra gli altri sono stati individuati i seguenti tool di monitoring:

- GridMap
- GridView
- Ganglia
- Gstat

2° settimana

La seconda settimana della mia permanenza a Ginevra, è stata dedicata per la maggior parte ad incontri con i responsabili dei progetti dei tool di monitoring individuati per valutarne pregi, difetti ed soddisfazione dei requisiti individuati.

L'idea utilizzata da GridMap è stata scelta come quella che maggiormente si adattava alle esigenze di molteplicità di viste e gestione della vista stessa. GridMap è basato su un nuovo uso di visualizzazione dell'algoritmo di TreeMap nel contesto del Grid. In GridMap, i siti ed i servizi nel Grid sono rappresentati come rettangoli di differenti dimensioni e colori, permettendo una visualizzazione bidimensionale dei dati. Questa rappresentazione dei dati monitorati richiede molto meno spazio che nelle convenzionali tabelle ordinate o nei grafici a barre. Cambiando tra le differenti viste GridMap permette di scoprire le correlazioni e i pattern associati fra i dati monitorati.

Una volta scelto di utilizzare l'idea di GridMap sono stati fatti ulteriori incontri con i membri del team e si è stato raggiunto un accordo per una collaborazione fra il progetto D4Science ed il progetto GridMap stesso. Nell'ambito del meeting tecnico D4Science, che si terrà a Ginevra nella prima settimana di dicembre, è previsto un ulteriore incontro tra i due team.

Se è quindi iniziato a progettare le estensioni all'esistente tool di monitoring sfruttando l'idea di base di GridMap. Il pre-esistente tool permetteva il browsing delle risorse ed una vista geo-referenziata dei diversi tipi di risorse gestite da D4Science.

Sono state quindi analizzate le modalità con le quali le nuove estensioni potevano essere integrate con il pre-esistente tool. A tal fine sono stati individuati i parametri che devono essere tenuti in considerazione nel passaggio da una vista all'altra, quali ad esempio la Virtual Organization (VO) da monitorare.

Per quanto riguarda l'estensione è stato effettuato uno studio su quali dovevano essere i parametri sui quali basare la vista a rettangoli. In particolare sono emerse le seguenti necessità:

- selezione della VO da monitorare
- 3 sotto viste da monitorare che rispecchiano lo stato di 3 tipi di risorse diverse gestite da D4Science: GHNs, Services, Data

A comune fra le 3 sotto viste sono stati selezionati 6 algoritmi noti in letteratura per il calcolo del TreeMap:

- Squarified
- Slice
- Strip
- Pivot
- Pivot Size
- Pivot Middle

3° settimana

Nell'ultima settimana ho cominciato l'implementazione prototipale delle estensioni previste.

Il problema è stato approcciato secondo il modello Control View Model. Questo modello permette di raggiungere il duplice scopo del tool di monitoring: avere una “vista” machine-oriented che può essere sfruttato dalle macchine per monitorare “automaticamente” l'infrastruttura; fornire una vista human-oriented.

Per quanto riguarda la parte human-oriented (interfacce grafiche), come toolkit di sviluppo è stato scelto il Google Web Toolkit (GWT) per i vantaggi di interattività ed astrazione dalle differenze fra i browser che fornisce. Questo permette allo sviluppatore di concentrarsi solo sulle funzionalità senza pensare ai grandi problemi di portabilità che sono valutati come l'80% dell'effort per uno sviluppatore di interfacce WEB.

Una volta partita l'implementazione è stato necessario definire le modalità di raggruppamento delle informazioni, i parametri di dimensionamento di primo e secondo livello dei rettangoli e infine gli stati in cui ogni rettangolo si può trovare.

Nell'ambito della vista sui GHN, come primo livello si rappresentano i site, mentre come secondo livello, si rappresentano GHN. In particolare fermi restando gli algoritmi per il calcolo del TreeMap il dimensionamento dei rettangoli (sia nel primo che nel secondo livello) può avvenire secondo i seguenti parametri:

- GHN
- RI
- CPU
- Bogomips
- Dimensione totale della memoria virtuale
- Dimensione della memoria virtuale disponibile
- Spazio disco

Gli stati in cui un rettangolo si potrà trovare sono:

- CERTIFIED
- SUSPENDED
- UNCERTIFIED
- UNREACHEABLE

Per quanto riguarda l'implementazione prototipale iniziata presso il CERN il mio lavoro ha visto affrontare solo questa vista.

Nel proseguimento della mia attività di ricerca è prevista la condivisione delle conoscenze e degli studi effettuati presso il CERN con altri membri del team di ricerca del quale faccio parte e che si affiancheranno a me nello sviluppo di questo tool di monitoring.

CONCLUSIONI

Il giudizio complessivo sull'esperienza presso il CERN è senz'altro positivo, sia per le numerose competenze acquisite sia per la crescita professionale che un periodo di studi di questo tipo comporta. L'approfondimento degli aspetti maggiormente legati a D4Science ed il generale scambio di competenze tra me l' Ing. Pedro Andrade e gli altri ricercatori del CERN credo che sia stato un buon risultato permesso da questo programma di Short Mobility.

Il rapporto di cordialità, ed in qualche caso anche di amicizia, stabilito con l'Ing Andrade faciliterà lo sviluppo ulteriore del monitoring tool nei prossimi mesi, e le estensioni sono già state adottate nell'ambito del progetto D4Science. Nel corso dei prossimi mesi avrò inoltre modo di trasferire le nuove conoscenze ai membri del consorzio D4Science in occasione dei periodici meeting tecnici tenuti a cui parteciperò.

Visto

Dott.ssa Donatella Castelli