Servizio Condor nell'INFN

presente e possibile futuro

Francesco Prelz
INFN, sezione di Milano

Sommario

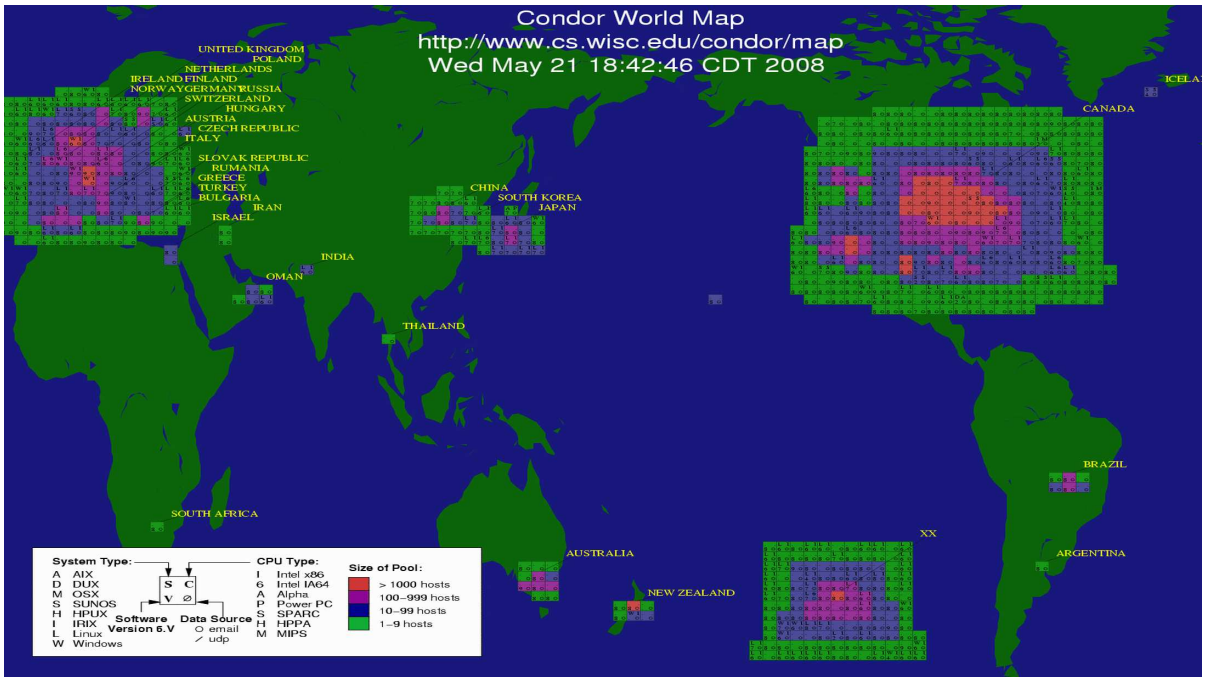
- Cos'è Condor
- Statistiche del pool INFN
- Il problema dell'accesso ai dati
- Possibili nuove applicazioni:
  - Interfaccia di accesso uniforme
  - Batch system "*free*"
  - Gestore di capacità "*back-fill*"
- Conclusioni

Cos'è Condor (1)

- *"A specialized workload management system for compute-intensive jobs"*
- Con alcune caratteristiche peculiari, come la capacità di utilizzare cicli di calcolo su infrastrutture primariamente utilizzate per altri scopi
  - ⇒ il punto di vista del *proprietario* della risorsa di calcolo non viene mai trascurato
  - Supporto nativo, da sempre nello "standard universe", di *checkpoint* e migrazione dei processi: *virtualizzazione delle applicazioni.*.
  - Supporto nativo del *VM universe*
- Progetto iniziato nel 1986 presso il dipartimento di Computer Science dell'Università del Wisconsin, Madison. Da allora Condor offre con risorse proprie circa 650 giorni di CPU al giorno agli utenti istituzionali di tutta l'Università: questo è il terreno di prova di ogni miglioria.
- Il principale punto di forza è la ricchezza semantica
  - della configurazione (e quindi della struttura dei singoli servizi)
  - del linguaggio usato per l'associazione dei job agli slot di esecuzione (*Classified Ads*)

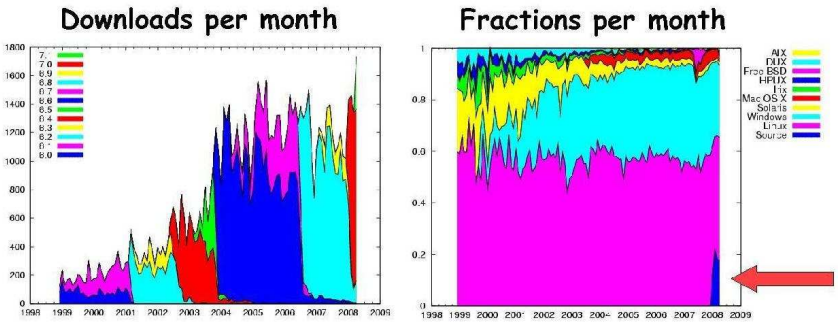
Scala attuale di Condor: pool nel mondo

- 850 pool, 145167 macchine.



Cos'è Condor (2)

- Un progetto *open source*, che, dopo varie traversie, è intenzionato a rimanere tale.

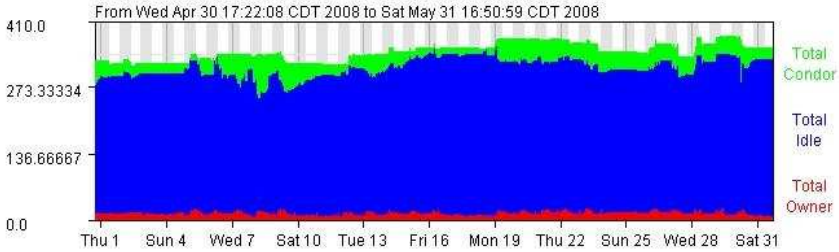


- Codice che potreste trovare pubblicizzato come parte di RedHat Enterprise MRG.
  - Dettagli in `http://www.cs.wisc.edu/condor/CondorWeek2008/condor_presentations/`
  - *"You might not know, but Condor is in Fedora, just yum install it to get the Condor 7.0.x series"* - per la "community grid" Fedora Nightlife.

Il pool INFN - statistiche recenti (1)

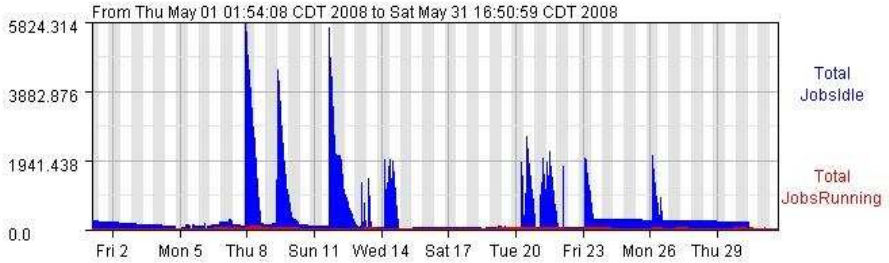
- `http://cmcondor.bo.infn.it/statistiche/`

- Nel corso dello scorso anno (Luglio 2007-Giugno 2008), il pool Condor nazionale
  - ha compreso una media di 321,4 processori (minimo 247, massimo 360), situati nelle sezioni di:
    - Bologna (83%), Pavia (10%), Milano (4%)
    - Milano Bicocca (3%), Padova e Torino (< 1%)
  - Media processori impegnati: 35,2 (11,0 % del totale)
  - Ore macchina allocate: 274250 (31,3 anni)
  - Media processori *idle*: 269,2 (83,7 % del totale)
  - Picchi mensili di processori impegnati: 92-184 (29 % - 57 %)



Il pool INFN - statistiche recenti (2)

- Esistono tuttavia frequenti picchi di 500 - 10000 job in coda.



  - Job del gruppo Alice di Bologna limitati a poche macchine sulla rete di Bologna.
  - Utenti totali: 23
    - Bologna 65%, Pavia 17%, Padova 13%, Torino 4%
    - Come dare loro un servizio *migliore* ?
  - FTE impegnati del gruppo di supporto: 0.6
  - Giorni di downtime complessivo negli ultimi 5 anni: 2

Il problema dell'accesso ai dati

- L'esistenza di job *data-intensive* non è nè ignorata nè trascurata: è stato il primo problema serio incontrato nel pool Condor INFN nel 1997!
- Lo "standard universe" di Condor prevede (e prevedeva già allora) l'accesso remoto ai dati, con cache locale configurabile.
- Chirp e Parrot/PFS sono la generalizzazione di questo modello.
  - `parrot vi /http/ccrws08.lngs.infn.it/`
  - `parrot bash - cat /http/ccrws08.lngs.infn.it/`

- 10 anni di ricerca e sperimentazione nel campo hanno portato:
  - Nella pratica, a fare in modo che i dati verso cui è richiesto accesso intensivo siano sempre presenti su dischi *montati localmente*
  - Nella teoria, a riconoscere che lo scheduling di operazioni relative ai dati (allocazione, trasferimento, cancellazione, deallocazione) ha la stessa dignità dello scheduling dei job.
- Stork, è un "Data Placement Scheduler". Permette di comporre DAG con operazioni sui dati.

Applicazione: interfaccia di accesso uniforme

- Negli ultimi anni è stato aggiunto un nuovo universo di sottomissione ("grid" universe) che consente di trasferire job ad altri sistemi:
  - Globus GT2/GT4 (Condor-G)
  - Unicore
  - Nordugrid
  - Batch system (BLAH - Sviluppo INFN-MI)
  - Altri pool Condor (Condor-C)
- Una installazione di Condor come puro "submitter node" consente di ottenere accesso uniforme a tutti questi sistemi. Altri canali di accesso (ad esempio verso CREAM) sono in via di sviluppo.
- Una installazione di Condor completa (detta anche "personal Condor") consente di sottomettere job *pilota* ("glide in") che aggiungono le risorse via via disponibili ad un pool accessibile attraverso i normali meccanismi di *match-making*.

Applicazione: batch system "*free*"

- Condor è un batch system "free", open-source e con una attiva comunità di supporto. Questa è la funzione svolta nella maggior parte dei pool Condor nel mondo, spesso, in ambiente industriale come frutto di analisi costi/benefici.
- Molte delle attività di sviluppo attualmente in corso sono volte a migliorare questa funzione su richiesta degli utenti:
  - Algoritmo di matchmaking specificabile dell'utente (jobrouter).
  - Policy di scheduling aggiuntive.
  - Hook per la connessione con altri sistemi.
- Circa un terzo del Tier-2 Atlas di Milano è stato sperimentalmente installato con questo sistema: i risultati della sperimentazione saranno disponibili a Ottobre.
- Provare ad abilitare il VM Universe ?

Applicazione: gestore di "*back-fill*"

- La peculiarità di Condor è sapere utilizzare i cicli di CPU disponibili con rispetto per le altre applicazioni, qualunque esse siano.

- Nodi di lavoro utilizzati in batch system esistenti potrebbero fornire la loro capacità residua al pool Condor nazionale.
- Esiste supporto in Condor anche per utilizzare la capacità di backfill disponibile per applicazioni BOINC (i vari progetti @HOME).
- Viceversa, se non c'è backfill da eseguire, comparirà presto anche supporto per il cosiddetto "*green computing*", per sospendere i nodi non utilizzati e risvegliarli all'arrivo di nuovo lavoro.

Conclusioni

- Il pool Condor INFN serve una comunità piccola ma fedele di utenti di applicazioni "personali", non inquadrate in framework più grandi.
  - Principalmente applicazioni in cui l'accesso a dati remoti non è l'attività principale, e l'investimento per passare a sistemi più complessi non è remunerativo.
- La capacità presente consente di gestire picchi di richieste e anche altri utenti di questa categoria. Ogni capacità aggiuntiva è utile a migliorare la risposta di picco, che è la caratteristica qualificante di questo sistema.
  - Se Fedora Nightlife decollasse, potremmo pensare non solo di contribuire a raggiungere il milione di nodi previsto , ma di utilizzare parte di questa capacità.
- Il mantenimento di un nucleo di competenza INFN su questo sistema in costante evoluzione consente di esplorare altre applicazioni:
  - Condor può utilmente servire come batch system "free".
  - Condor consente un accesso flessibile ad altri sistemi di calcolo distribuito ed il trasferimento di job anche in modalità di *back-fill*.
  - Etc. etc.

Grazie!

- *Grazie* per l'attenzione e per aver resistito alla tentazione di tornare a casa prima!
- Per domande e discussioni: `condor@infn.it`, `condor-users@cs.wisc.edu`
- Se volete installare Condor:
  - nella versione stabile attualmente supportata (6.8.5) con la configurazione pronta per partecipare al pool INFN: `http://www.bo.infn.it/calcolo/condor/HA/infn-installation-tool.htm`
  - Per accedere a tutte le versioni attive, a tutte le piattaforme supportate, e ai sorgenti: `http://www.cs.wisc.edu/condor/downloads-v2/` e contattate `condor-admin@infn.it` per avere aiuto per la configurazione.
  - Altrimenti, se avete Fedora Core 9, provate `yum install condor` Attenzione però: per problemi di licenza l'RPM in Fedora non include per il momento lo standard universe (glibc modificata: il problema dovrebbe essere risolto presto), e la sottomissione a Globus e Nordugrid.