



GESTIONE DELLE CATENE OPERATIVE PREVISIONALI NEL SETTORE METEO - IDROLOGICO AL CINECA

GIAN FRANCO MARRAS

Costituito nel 1969 (come Consorzio Interuniversitario per il Calcolo Automatico dell'Italia Nord Orientale), oggi il CINECA è il maggiore centro di calcolo in Italia e uno dei più importanti a livello mondiale.

CINECA è un Consorzio Interuniversitario senza scopo di lucro formato da 70 Università italiane, 5 Enti di Ricerca Nazionali (tra cui CNR e OGS4) e il Ministero dell'Istruzione e della Ricerca Scientifica (MIUR). Operando sotto il controllo del MIUR, offre supporto alle attività della comunità scientifica tramite il supercalcolo e le sue applicazioni, realizza sistemi gestionali per le amministrazioni universitarie e il MIUR, progetta e sviluppa sistemi informativi per pubblica amministrazione, sanità e imprese.

Sempre più punto di riferimento unico in Italia

Sopra: "Cross session vertical velocity + volume rendering specific humidity". Immagine realizzata in collaborazione con ArpaE e prodotta con il software "Vapor" (Computational Information Systems Laboratory at the National Center for Atmospheric Research)

per l'innovazione tecnologica, con sedi a Bologna, Milano e Roma e oltre 700 dipendenti, il CINECA opera al servizio di tutto il sistema accademico, dell'istruzione e della ricerca nazionale, attraverso il calcolo scientifico ad alte prestazioni, la gestione e lo sviluppo di reti e servizi telematici, la realizzazione di sistemi informativi articolati e complessi per il trattamento di grandi quantità di dati.

L'infrastruttura per il supercalcolo e le competenze di CINECA in questo settore sono all'avanguardia per quanto riguarda tutti gli aspetti, dal calcolo ad alte prestazioni alla visualizzazione scientifica, dalla gestione dello storage all'infrastruttura di rete.

Il CINECA nasce come strumento per fornire alla ricerca nazionale strumenti di calcolo particolarmente avanzati e costosi, nell'ambito del calcolo

ad alte prestazioni (HPC) o supercalcolo.

La sua missione si è mantenuta nel tempo e oggi CINECA dispone della più grande infrastruttura HPC italiana, a disposizione della ricerca accademica, della ricerca pubblica e industriale, che rende possibili simulazioni ed esperimenti computazionali altrimenti non perseguibili con risorse convenzionali.

Dal settembre del 1993, il CINECA è stretto collaboratore con Arpae nelle attività inerenti il servizio di previsione meteorologica e marina nazionale. Infatti il CINECA nel tempo si è aggiudicato diversi bandi comunitari per l'affidamento di servizi e fornitura di risorse di supercalcolo per la modellistica previsionale meteorologica e marina indetti dal servizio Idro-Meteo-Clima (Arpae-SIMC) di Arpae. Grazie alla collaborazione con il CINECA OGS produce i dati relativi alle analisi biogeochimiche sul mare Mediterraneo grazie al progetto "Copernicus: marine environment monitoring service", un servizio che a livello europeo fornisce informazioni sulla qualità del mare e sulle condizioni ambientali dei mari ed oceani che costeggiano il continente.

Grazie a queste collaborazioni, oltre che con Arpae-SIMC e OGS, anche con Arpa-Piemonte, CNR ed altri ancora, il CINECA ha acquisito numerose competenze sulla modellistica previsionale sia in ambito meteorologico e meteo-marino che relativamente al controllo dello stato della qualità dell'aria e dei mari.

Ad oggi CINECA può vantare competenze su quattro fronti distinti:

- gestione e amministrazione di sistemi computazionali complessi come quelli del supercalcolo;
- gestione di codici in ambiente supercalcolo: parallelizzazione, ottimizzazione e debugging;
- conoscenza dei modelli previsionali dal punto di vista informatico per quanto riguarda la loro configurazione e il loro corretto funzionamento;
- gestione di servizi ad elevata criticità in ambienti di supercalcolo.

Riguardo al primo punto il CINECA è leader in Italia ed è uno dei centri di eccellenza a livello mondiale. La gestione dei propri sistemi HPC richiede specifiche competenze di tipo sistemistico, legate alla dimensione dei sistemi e all'utilizzo di tecnologie di ultima generazione. Il personale CINECA è altamente specializzato a tali esigenze, grazie alla sua storia quasi cinquantennale nel campo.

Per quanto riguarda il secondo e terzo punto,

il personale CINECA svolge di norma attività di parallelizzazione, ottimizzazione e debugging di codici scientifici in collaborazione con vari enti di ricerca italiani ed europei. In particolare ha specifica competenza sulle tecniche numeriche e di parallelizzazione di modelli meteorologici, climatologici e oceanografici. Varie componenti dei codici COSMO, SWAN, Farm e molti altri ancora sono state direttamente corrette e ottimizzate dal personale CINECA.

La tipologia di servizio che solitamente viene richiesto nel campo meteo - idrologico accoppia le tipiche problematiche dei sistemi di supercalcolo con quelle dei servizi ad elevata criticità. Normalmente questi due tipi di servizi sono disgiunti. L'operatività di un servizio previsionale che utilizza il supercalcolo, invece, deve funzionare in modalità automatica 365 giorni all'anno con tempistiche rigorose e con un'aspettativa di affidabilità e infallibilità notevolmente elevata. Per erogare un servizio di questo genere, ogni parte dell'infrastruttura deve essere estremamente affidabile e ridondata in ogni sua parte e l'ambiente deve essere stabile nel tempo e resistere ad ogni evento critico che può compromettere la sua funzionalità.

GESTIONE DEI SERVIZI

I servizi attualmente erogati dal CINECA si basano su una efficiente e robusta infrastruttura tecnologica ed operativa che, per mantenere un alto livello di affidabilità, ha impostato soluzioni specifiche.

I servizi previsionali si basano su una catena operativa, un flusso di comandi completamente automatizzato dove l'intervento dell'uomo risulta occasionale e necessario solo in casi di alta criticità. Tale flussi, comunemente chiamati workflow, sono stati sviluppati interamente da CINECA; sono formati da una serie di job suddivisi a loro volta da vari jobstep. Ogni jobstep esegue una serie di comandi che gestiscono: i) la pre-elaborazione dei dati di input; ii) l'esecuzione dei modelli previsionali; iii) la post-elaborazione dei dati di output; iv) la loro archiviazione per l'elaborazione successiva. Al singolo workflow è associato un sistema di monitoraggio delle varie fasi e per ogni run eseguito viene prodotto un resoconto finale o report.

Le catene operative, operano in ambiente UNIX e, nel tempo, si sono perfettamente integrate con architetture di tipo diverso (generiche distribuzioni linux e sistemi IBM AIX). Per la sottomissione dei job si utilizza nei nostri sistemi lo



Vista della nuova sede del Cineca e "Marconi", il nuovo supercomputer a disposizione della ricerca Italiana ed Europea

scheduler PBS, ma sono supportati anche Load Leveler e LSF ed è possibile sviluppare backend specifici verso altri scheduler.

Le catene operative hanno una gestione automatica degli incidenti e risolvono in completa autonomia gran parte dei problemi che si possono presentare. Per esempio individuano e riservano le risorse necessarie, verificando che siano funzionanti. In caso di non completa disponibilità possono partire in modalità degradata e, grazie al grado di ridondanza assicurato, sono di norma in grado di portare a termine il servizio nei tempi richiesti.

Altre situazioni di errore gestite in automatico sono relative al workflow interno della parte modellistica. Normalmente le catene operative devono acquisire preliminarmente i dati di input, sia

scaricandoli dai server preposti che accedendo ai dati prodotti dal run precedente. Se quel run non era finito correttamente, potrebbe esservi un problema di correttezza del run successivo, ma nel nostro workflow questa situazione viene recuperata, prima cercando i dati mancanti dal sistema di backup, poi facendo partire comunque la corsa in modalità degradata.

Il sistema di backup, nella attuale configurazione, è un secondo sistema di calcolo che esegue, a caldo (hot-backup), le stesse operazioni del sistema principale, in una cosiddetta modalità "mirroring". Le varie istanze delle catene installate sui diversi sistemi sono del tutto indipendenti, possono operare simultaneamente garantendo un'alta affidabilità del servizio. La modalità mirroring è attivabile anche solo per specifici periodi, per esempio nei fine settimana e nei periodi festivi per sopperire alla diminuita disponibilità dell'operatore. Se l'operatore lo ritiene opportuno, tale sistema può essere attivato in tempi relativamente brevi. Infatti, anche se il backup non è attivo, i dati vengono comunque predisposti per minimizzare i tempi di attivazione.

Tutte le istanze della catena operativa sono eseguite in modalità automatica tramite crontab con una pianificazione periodica dei comandi necessari al servizio. Ogni sistema è configurato con dispositivi e servizi ridondati per garantire High Availability (HA), cioè un'alta disponibilità e affidabilità del servizio stesso. Per esempio, ogni sistema ha più nodi di login e in caso di fallimento di uno, l'altro è in grado di sostituirsi in modo automatico. Tutti i comandi necessari per l'esecuzione delle catene operative sono registrati sul crontab di almeno due nodi di login. Un sistema di controllo gestisce il corretto funzionamento di ogni singola procedura registrata evitando eventuali conflitti fra i nodi.

Per garantire maggiore stabilità e performance, sul sistema principale le catene utilizzano un filesystem dedicato e separato da quello del resto dell'utenza. Sui sistemi di backup viene utilizzato un filesystem che, pur non essendo dedicato, ha comunque una affidabilità superiore a quello standard. Su entrambi i sistemi è comunque possibile utilizzare un secondo filesystem in tempi molto brevi in caso di failure del primo.

SVILUPPO E MESSA IN PRODUZIONE DELLA CATENA OPERATIVA

Particolare impegno è stato rivolto agli aspetti di sviluppo e di messa in produzione delle catene operative. Si tratta di un processo

complesso per varie ragioni:

- le catene vengono costantemente aggiornate; alcune modifiche vengono realizzate su specifica richiesta del Committente, altre vengono introdotte direttamente da CINECA al fine di migliorare ed irrobustire il servizio;
- qualsiasi modifica deve essere testata adeguatamente prima di essere messa in produzione;
- il rilascio in produzione di una nuova release delle catene deve avvenire in maniera schedulata, secondo regole ben precise. Solo in caso di emergenza deve essere possibile applicare modifiche immediate alla catena operativa (ad esempio per evitare la perdita delle corse immediatamente successive);
- L'attivazione di un sistema di backup o la sostituzione del filesystem deve essere estremamente rapida;
- Tutte le istanze su macchine diverse, devono essere mantenute allineate, cioè devono fornire risultati equivalenti dal punto di vista modellistico;
- in qualsiasi momento deve essere possibile conoscere esattamente le condizioni sotto le quali è stata eseguita una corsa operativa, ed eventualmente eseguirla nuovamente con le stesse identiche modalità.

La singola catena operativa è un'entità unica costituita da:

- sorgenti dei programmi utilizzati (modelli meteorologici, interpolatore dei dati, strumenti accessori);
- strumenti per compilare ed installare i sorgenti;
- script per la gestione del workflow;
- file di configurazione del servizio;
- file di input "statici", ovvero che non cambiano da un run all'altro.
- file di input "dinamici", ovvero che cambiano da un run all'altro.

Le catene sono sotto il controllo di un sistema di versioning (SVN) per facilitarne lo sviluppo e la manutenzione. È definito un preciso protocollo, che ha come obiettivo la razionalizzazione del processo di messa in produzione delle modifiche. Oltre ai pacchetti dei sorgenti, nel sistema di controllo di versione troviamo anche i compilati degli eseguibili e delle librerie associate. Questo rende più rapido il processo di passaggio della catena in ambiente diverso senza il bisogno di ricompilare tutti i sorgenti. Inoltre si hanno a disposizione più versioni dello stesso pacchetto in modo tale da passare in tempi rapi-

di da una versione all'altra. Questo caso è stato utile in passato, per esempio, quando abbiamo riscontrato un errore della versione in uso che non era presente nella versione precedente.

Dovendo gestire più istanze su macchine diverse, si è fatta la scelta di minimizzare la duplicazione di codice. La catena è un'unica entità, funzionante su tutte le macchine previste. La gestione delle differenze viene fatta attraverso librerie di astrazione che forniscono una interfaccia comune (front-end), ed uno o più back-end specifici della piattaforma, che tengono conto delle differenze in termini di scheduler e di sistemi operativi. I dati di input vengono acquisiti principalmente da un server preposto e in alcuni casi tramite una connessione dedicata; in caso di irraggiungibilità, i dati vengono cercati su un server secondario. La catena operativa gestisce entrambi i server e passa automaticamente da uno all'altro in caso di problemi. La distribuzione dei dati prodotti dalla catena operativa avviene attraverso tre modalità principali:

- un sottoinsieme dei dati viene caricato sul server del Committente;
- i dati sono disponibili su un server ftp presso CINECA;
- i dati possono essere visualizzati, attraverso un sito web che pubblica le mappe relative ai principali campi previsionali.

Sia il server ftp che il sito web sono protetti da credenziali (username e password) assegnate ad ogni singolo utente. Questi servizi sono ospitati su un cluster di servizio, che si occupa di raccogliere i dati prodotti dai sistemi di supercalcolo. L'archiviazione dei dati viene eseguita principalmente su nastro; i dati vengono salvati su disco in una particolare area di archiviazione dalla catena operativa e successivamente migrati su nastro. Alcuni dati vengono salvati per sempre, mentre altri, che potrebbero essere ricreati in poche ore, vengono salvati solo per 7 giorni.

CONCLUSIONI

La tipologia di servizio che solitamente viene richiesto nel campo previsionale deve essere sempre garantita e i risultati devono essere disponibili in tempi certi. La gestione delle catene pertanto necessita di sistemi di calcolo ridondati in ogni singola componente, un workflow completamente automatizzato che risolve in piena autonomia i problemi più disparati e personale operativo altamente qualificato in più fronti, dal punto di vista sistemistico e informatico.

Gian Franco Marras