



Progetto di Interesse strategico NEXTDATA

Rendicontazione scientifica per il periodo di riferimento **01/01/2012-31/12/2012**

Unità CASPUR

WP 2.5 - Archivio digitale di dati numerici e previsionali

1. Attività prevista e risultati attesi (come indicato sul Piano Esecutivo, inclusi i milestones)

La disponibilità di risultati prodotti da simulazioni numeriche climatiche, paleoclimatiche e ambientali è un elemento essenziale per completare e interpretare le informazioni fornite dai dati misurati e sviluppare metodologie di previsione e stima degli impatti. Le attività dell'Unità CASPUR contribuiranno alla costruzione di un archivio dei risultati di simulazioni climatiche globali e regionali, sia già esistenti sia effettuate appositamente durante il progetto.

A livello di condizioni climatiche globali, in collaborazione con l'Unità CNR-ISAC verranno costruiti archivi dei risultati delle simulazioni relative al periodo industriale (1850-2005) e a proiezioni future (RCP 4.5, RCP 8.5, RCP 3-PD) per il periodo 2006-2100, ottenute utilizzando il modello globale EC-Earth installato presso il CASPUR. Le simulazioni globali saranno in parte derivate da simulazioni già realizzate ed in parte provenienti da nuove simulazioni di EC-Earth che saranno realizzate sulle macchine del CASPUR nel corso del primo anno. CASPUR fornirà anche, ove possibile, supporto all'ottimizzazione dei codici numerici.

In collaborazione con le Unità ICTP e CNR-ISAC, verrà data disponibilità alla realizzazione da parte di ICTP di simulazioni regionali per le aree di interesse del progetto con il modello RegCM eventualmente innestato in EC-Earth, che verranno incluse nell'archivio di dati numerici. L'archivio comprenderà anche una casistica dei risultati di simulazioni non-idrostatiche ad alta risoluzione (1-10 km) in aree montane, in cui l'orografia complessa richiederà lo sviluppo di particolari soluzioni modellistiche, realizzate da CNR-ISAC e da altre strutture, in parte sulle macchine del CASPUR, con il modello WRF. In queste aree, il confronto fra i risultati numerici e i dati raccolti dalle reti di misura sarà essenziale per la calibrazione e la validazione dei modelli a scala locale. Questi modelli saranno innestati nelle simulazioni globali per ottenere un archivio di scenari futuri ad alta risoluzione spaziale nelle regioni montane di interesse. CASPUR fornirà anche, ove possibile, supporto all'ottimizzazione dei codici numerici.

Primo anno:

Partecipazione al censimento dei risultati delle simulazioni e delle rianalisi numeriche globali e regionali disponibili presso i partecipanti al progetto, all'armonizzazione dei protocolli di archiviazione dei dati numerici e delle modalità di accesso ai dati e delle procedure di trasferimento di dati di grandi dimensioni e inizio delle attività di archiviazione dei risultati numerici disponibili.

Partecipazione alla definizione e preparazione di specifici esperimenti numerici da svolgere nel corso del progetto e ottenimento dei primi risultati di nuove simulazioni globali e regionali per le aree di interesse del progetto.

Entro la fine del primo anno, in collaborazione con le Unità CMCC, CNR-ISAC, ENEA UTMEA e ICTP, CASPUR darà un contributo alla realizzazione della prima versione dell'archivio numerico.

2. Deliverables previsti per il periodo di riferimento

Contributo al D2.5.1 (PM12): Relazione sul censimento delle simulazioni climatiche.

3. Attività effettivamente svolta durante il periodo di riferimento

3.1 Attività di ricerca

Nessuna

3.2 Sviluppi applicativi, tecnologici e informatici

CASPUR ha messo a disposizione tutta la tecnologia disponibile per il calcolo ad alte prestazioni, *storage* ed elaborazione dati necessaria alle applicazioni di frontiera del presente progetto.

L'infrastruttura utilizzata per NextData comprende un tipico supercalcolatore a memoria distribuita basato su processori tradizionali X86 e un sistema di *server* per il servizio di I/O intensivo (e per i servizi di accesso esterno, virtualizzazione e DBMS) i quali sono interfacciati ad un sottosistema di *storage* di classe *Enterprise*.

Schematicamente viene riportato il diagramma a blocchi dell'intera infrastruttura *hardware*(Figura 1):

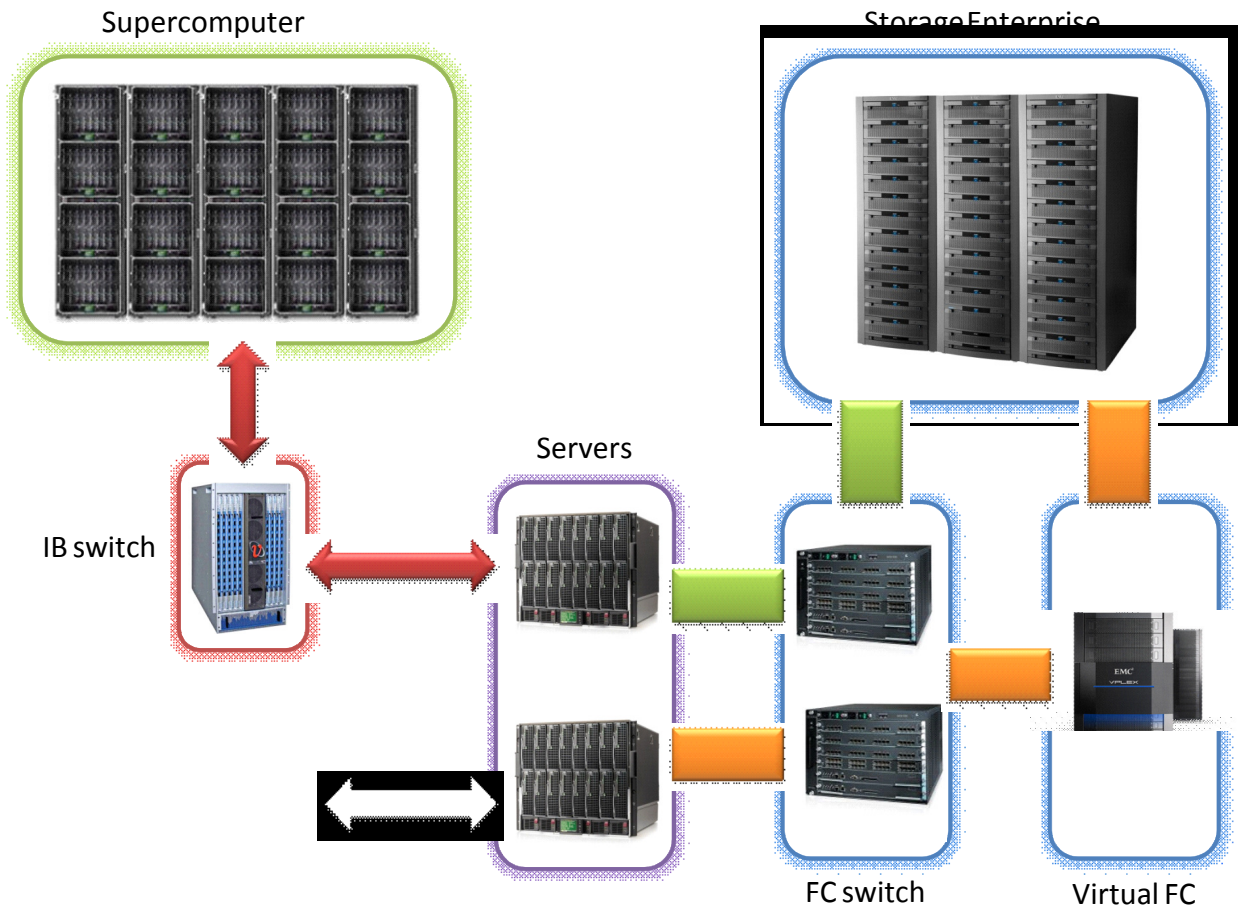


Figura 1. Diagramma a blocchi dell'infrastruttura ICT/HPC per NextData.

Questa infrastruttura di calcolo e di *storage* ad altissime prestazioni ha consentito di supportare tutte le simulazioni previste a CASPUR e di iniziare ad immagazzinare i dati provenienti da simulazioni realizzate altrove. Il trasferimento della gran mole di dati disponibili altrove ha fortemente beneficiato della rete della ricerca GARR, rete della quale CASPUR è uno dei nodi strategici. Questa sinergia tra CASPUR e GARR ha garantito e garantirà agli utenti NextData una ampiezza di *banda* disponibile e un supporto specialistico e continuativo per la risoluzione di problematiche legate ad interruzioni, *bottlenecks* e trasferimento di dati massivi (dell'ordine di *terabytes* di dati).

Per rendere successivamente accessibili i dati computazionali censiti e/o prodotti all'interno del progetto NextData all'intera comunità scientifica interessata si è deciso di utilizzare il THREDDS Data Server (TDS) versione 4.2, aggiornato successivamente alla versione 4.3, dove THREDDS è acronimo di *Thematic Real-time Environmental Distributed Data Services*.

Il TDS è un web server che fornisce accesso a dati e metadati di tipo scientifico, usando differenti protocolli per accedere i dati remoti, tra i quali OPeNDAP, OGC WMS e WCS, HTTP. Thredds data server si appoggia al web server open source Apache Tomcat, che è una implementazione delle tecnologie Java Servlet e JavaServer Pages. Il modo in cui lavora il TDS è rappresentato nella figura 2.

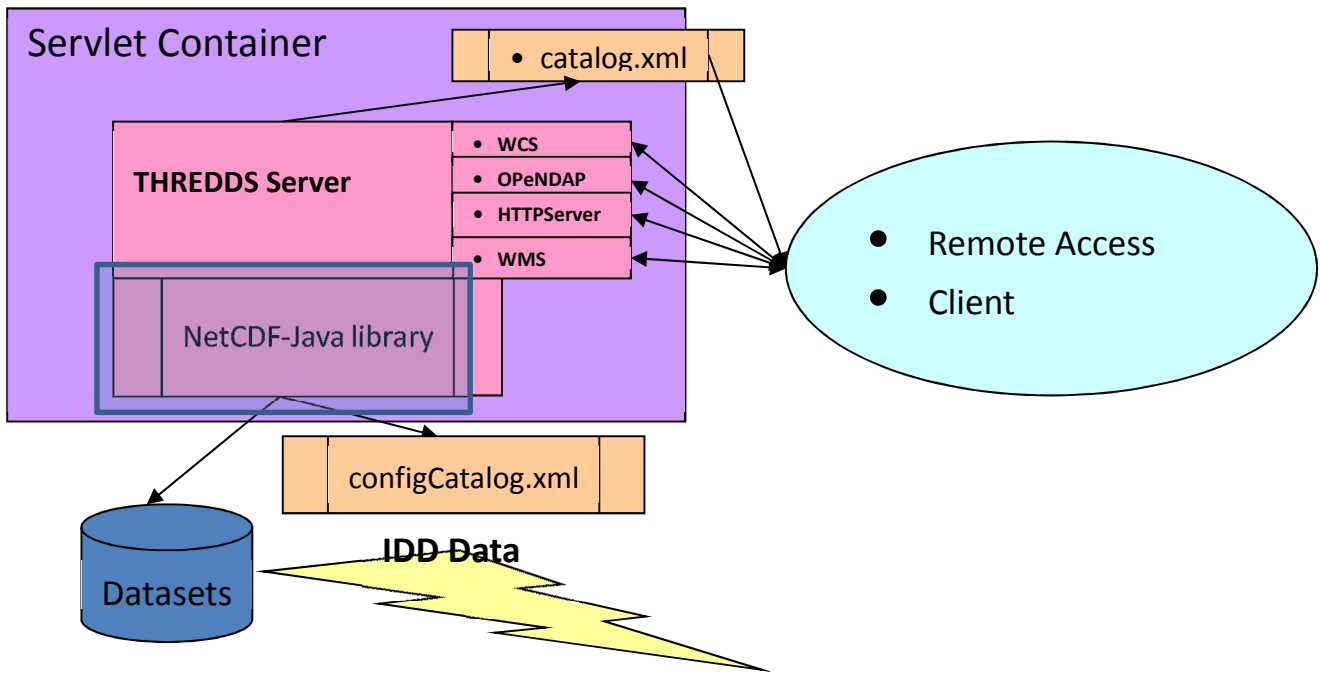
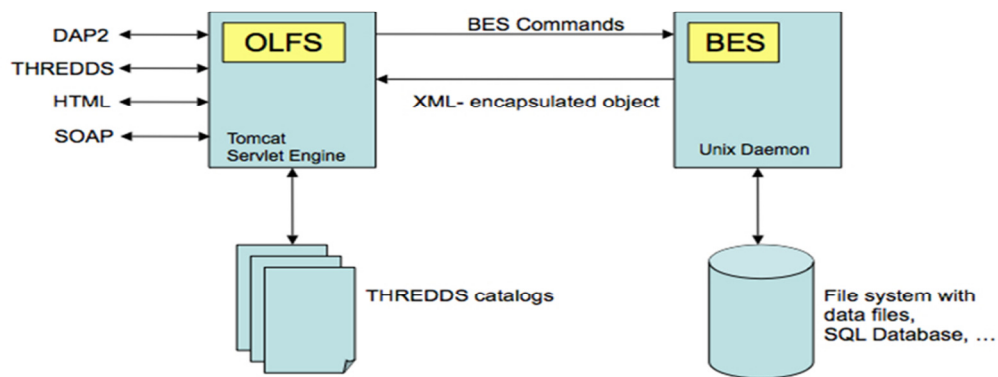


Figura 2. Schema del server TDS.

Al CASPUR è stato installato il TDS 4.2.10 (aggiornato successivamente alla 4.3) con protocollo OPeNDAP; la versione utilizzata di Apache Tomcat è la 7.0.30 (aggiornato alla 7.0.32) e di Java la 1.6. La macchina su cui è attualmente installato il TDS è bl102.caspur.it, che è una macchina con 4 processori Dual-Core AMD Opteron a 2,8 Ghz. I dati sono, invece, materialmente presenti sul Supercomputer CPU e resi disponibili da un *file system* di tipo *Lustre*, che permette di accedere da diversi computer ai file contenuti su un computer remoto tramite rete, potenzialmente in simultanea. Più precisamente *Lustre* è un *file system* parallelo distribuito ad alte prestazioni in grado di supportare reti di *cluster* con migliaia di nodi e *petabyte* di dati da immagazzinare. Il *file system* risiede sul sistema di *storage Enterprise* di capacità complessiva superiore a 1PB descritto in precedenza. In particolare è stata implementata la struttura descritta nella figura seguente, tenendo presente che oltre 60 TB di dati sono inoltre disponibili sul sistema di storage a nastro.

Hyrax Architecture



OPeNDAP Developer's Workshop Feb 21-23 2007

Hyrax è il nome alternativo per OPeNDAP Data Server

- OLFS (OPeNDAP Lightweight Front-End Servlet)
- BES (Back-End Server)

3.3 Attività di formazione

Nessuna.

3.4 Attività di disseminazione e divulgazione

Nessuna.

3.5 Partecipazione a conferenze

Non sono state previste partecipazioni a conferenze nel periodo.

4. Risultati ottenuti durante il periodo di riferimento

4.1 Risultati specifici (banche dati, risultati delle misure, output di modelli, etc)

Nessuna.

4.2 Pubblicazioni

Nessuna.

4.3 Disponibilità di dati e output modellistici (formato, supporto, etc)

Il Thredds Data Server del CASPUR ospita al momento i dati di alcune simulazioni ICTP e CNR-ISAC.

4.4 Deliverables completati

Il censimento dei dati provenienti da simulazioni climatiche è in corso.

5. Commento su eventuali scostamenti fra attività/risultati/deliverables previsti ed effettivamente realizzati

Nessuno.

6. Attività previste per il periodo successivo

A regime, il Thredds Data Server del CASPUR dovrà consentire un accesso semplice, standardizzato (secondo determinati criteri) ed efficiente ai dati del progetto NextData, siano questi ultimi presenti fisicamente a Roma (presso il CASPUR) o altrove, prodotti e/o censiti dai gruppi di ricerca del CNR-ISAC (Torino) e ICTP (Trieste) partecipanti al progetto.