

Minute della riunione con ARPA tenuta in via telematica il 12 aprile 2021 alle 14:30

Si esaminano inizialmente i problemi incontrati da ARPA durante l'utilizzo di ReCaS-Bari.

Angela Morabito comunica che nell'ultimo week-end le catene modellistiche previsionali non sono andate a buon fine, pertanto non è stato possibile produrre le mappe giornaliere previsionali sulla qualità dell'aria che vengono pubblicate ed aggiornate con cadenza giornaliera sul sito <http://cloud.arpa.puglia.it/previsioniqualitadellaria/>.

Premesso che questo non deve succedere mai, perché la pubblicazione delle mappe è un compito istituzionale di ARPA, alla domanda: "quanto spesso si verifica questo inconveniente" è stato risposto che non si verifica di frequente, anzi è piuttosto raro. Di solito se si vede che il workflow non riesce a girare, viene contattato Giacinto Donvito che sistema la cosa. In questo caso però, poiché la mancanza dei forecast si è verificata nel week-end, non è stato possibile informare tempestivamente Donvito. I run previsionali inizializzati lunedì 12 aprile 2021 sono andati a buon fine anche se con un pò di ritardo, adesso è necessario girare il workflow sui dati di domenica 11 e sabato 10 aprile.

Dalla discussione emerge che sarebbe opportuno disporre di un sensore che qualora la procedura della produzione delle mappe non parte lancia un allarme da inviare sia ad ARPA che a ReCaS-Bari.

Viene inoltre riferito che è stato di recente riscontrata una lentezza nell'uso del file system. Giacinto informa che questo è un problema comune a tutti gli utenti, causato da un problema hardware al sistema di storage.

Viene anche chiesto se convenga utilizzare *lustre* oppure *lustrehome*.

Donvito afferma che *lustre* è più veloce by design, ma ha meno protezioni: *lustre* è a singola copia, mentre *lustrehome* fa una doppia copia dei dati e questo lo fa diventare più lento.

Vengono anche chiesti dettagli circa il backup su nastro magnetico. Data la velocità di scrittura su nastro e la mole di dati da backuppare, si riesce a fare un backup completo di *lustrehome* ogni settimana. È possibile però accelerare il backup per alcuni file o cartelle ritenute più preziose. Per queste cartelle, se ritenuto necessario, si può arrivare a fare un backup giornaliero. È necessario perciò individuare quali siano i file che richiedono questo trattamento. I diversi utenti ARPA sono pertanto invitati a trasmettere a Donvito l'elenco dei file (o delle cartelle) che loro ritengono più preziosi insieme con la periodicità del backup (il passaggio attraverso Angela Morabito già tentato in passato, non è riuscito a concludersi perché sono diverse le applicazioni e gli utenti di Arpa Puglia su ReCaS, inoltre le attività che vengono svolte sono molto evolutive).

Viene anche chiesto quale è la policy per fare delle copie su nastro magnetico di dati stabili per preservarli per un lungo periodo. Giacinto spiega che al momento è necessario fare una richiesta con l'indicazione del file da preservare. In futuro l'utente potrà decidere

autonomamente se tenere un file solo su disco magnetico, su disco magnetico e nastro magnetico o solo su nastro magnetico.

È seguita una breve discussione per capire quali dati vanno preservati su nastro magnetico: i dati, provenienti dai modelli meteorologici o di qualità dell'aria a risoluzione maggiore o da vari database, subiscono diversi cambiamenti di formato per essere messi nella forma utilizzabile dalle catene modellistiche. Si è convenuto che i dati intermedi non sono necessari e che può essere utile conservare solo il set di dati che viene dato in pasto alle catene. È anche emerso che occorre preservare i file che definiscono la configurazione del modello, nel caso si voglia rilanciare la catena.

È stato annunciato che in ARPA sta maturando l'idea di effettuare le previsioni e le analisi di qualità dell'aria con una risoluzione spaziale di 1 km al posto dei 4 km attuali. ARPA a riguardo chiede quali nuove risorse hardware ReCaS-Bari possa mettere a disposizione per affrontare questo problema visto che i 40 core attualmente disponibili sui server HPC non saranno sufficienti. Donvito ha annunciato che i server del cluster HPC saranno sostituiti da server con più di 200 core, forse perfino 256 (dipenderà dall'esito della gara). I primi di questi server saranno disponibili in al massimo due mesi. Questo dovrebbe consentire di poter girare la gran parte dei job MPI all'interno del server, rendendo più veloce lo scambio di informazione tra un core e l'altro. Qualora poi fossero necessari più di 256 core, sarà sempre possibile implementare la configurazione attuale del cluster HPC, con interconnessioni tra i server con InfiniBand a 100 Gbps.

Donvito consiglia di verificare subito la scalabilità delle catene previsionali per capire qual è il numero massimo di core che può essere utilizzato.

Naturalmente l'obiettivo di ARPA è quello di andare ad una risoluzione sempre migliore. Si sta quindi valutando anche la possibilità di utilizzare immagini da satellite e, in particolare, i dati Copernicus e di implementare metodi statistici (Random Forest, ad esempio).

A questo proposito viene chiesto quale ruolo ha ReCaS-Bari nel progetto Mirror Copernicus. Donvito afferma che sta lavorando a questo progetto ormai da circa un anno e che, anche se niente è ancora definito, lui ha lavorato in maniera da avere Bari su tutta la linea, sia come repository di dati che come fornitore di servizi all'utenza finale

Donvito comunica inoltre che quasi tutti i server HPC saranno dotati di GPU. ARPA però non ritiene di poterle utilizzare, in quando la conversione dei programmi viene considerata troppo onerosa.

Si passa quindi alla presentazione delle nuove feature, alcune delle quali sono già disponibili, altre lo saranno a breve su ReCaS-Bari.