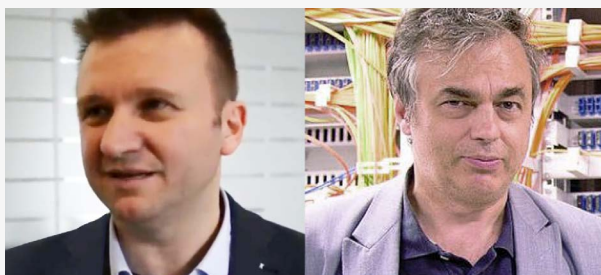


» INTERVISTA



Carlo Cavazzoni

Luca dell'Agnello

IL SUPERCALCOLO A SERVIZIO DELLA LOTTA CONTRO LA PANDEMIA

Intervista a Carlo Cavazzoni, ricercatore del consorzio interuniversitario CINECA di Bologna e a Luca dell'Agnello, ricercatore del centro di calcolo CNAF dell'INFN, di Bologna.

Il supercalcolatore Marconi del consorzio Cineca di Bologna, tra i più potenti del mondo, che sarà seguito nel 2021 dal supercalcolatore pre-exascale Leonardo, ha messo a disposizione dell'emergenza epidemica in corso le sue eccezionali prestazioni. In grado di eseguire decine di milioni di miliardi di calcoli al secondo, da febbraio è al lavoro per studiare il comportamento delle proteine che consentono al Covid-19 di replicarsi, al fine di accelerare la produzione di farmaci in grado di neutralizzare il virus. A fianco di Marconi in questa impresa c'è anche il CNAF, il Centro di Calcolo dell'INFN, che mette a disposizione, oltre alle sue infrastrutture di calcolo, anche la sua capacità di gestire i Big Data.

Fulcro della collaborazione è Exscalate, la piattaforma di supercalcolo con le più elevate prestazioni al mondo, progettata in collaborazione tra CINECA, l'azienda farmaceutica Dompé e il Politecnico di Milano per valutare 3 milioni di molecole al secondo, a partire da una "biblioteca chimica" di 500 miliardi di molecole, e in grado di operare contemporaneamente su 30 target biologici differenti.

Al fine di integrare al meglio le competenze necessarie ad accelerare e ottimizzare il processo, è stato istituito nei primi mesi dell'anno Exscalate4CoV, un consorzio guidato da Dompé, costituito da diversi istituti universitari e centri di ricerca, tra i quali l'INFN: il consorzio si è aggiudicato all'inizio di marzo un finanziamento di 3 milioni di euro stanziato con apposito bando della Commissione Europea (nell'ambito di Horizon 2020) destinato a progetti di ricerca sul Coronavirus.*

Obiettivo primario di Exscalate4CoV (E4C) è sfruttare le potenzialità di supercalcolo e delle piattaforme Big Data integrate con le migliori competenze scientifiche in ambito biomedico, per fronteggiare situazioni di pandemia di interesse sovranazionale. Oltre a selezionare le molecole più promettenti per contrastare l'attuale epidemia, infatti, il consorzio si è posto come obiettivo quello di dare forma a un modello operativo di intervento a livello europeo, efficace anche in futuro per eventi analoghi. I risultati prodotti dal consorzio saranno messi a disposizione della comunità scientifica.

Abbiamo chiesto a Carlo Cavazzoni, responsabile per CINECA della piattaforma Exscalate e a Luca Dell'Agnello, referente per INFN-CNAF del consorzio Exscalate4CoV, di raccontarci genesi e stato dell'arte del progetto.

» INTERVISTA

Dr. Cavazzoni, le chiederei intanto di presentarci il protagonista, il supercalcolatore Marconi? Per cosa è stato progettato? Partecipano al consorzio altri supercalcolatori europei?

[Carlo Cavazzoni, Cineca] Marconi è attualmente il principale supercalcolatore italiano inserito all'interno della network dei supercalcolatori europei chiamata PRACE (*Partnership for Advanced Computing in Europe*), in quanto a dimensione e potenza Marconi è nei top3 europei e nei top10 mondiali. Marconi è stato progettato principalmente per affrontare le grandi sfide scientifiche del nostro tempo, inclusa la fusione nucleare, la progettazione nuovi materiali, la scoperta di nuove particelle, lo studio dell'espansione dell'universo, i cambiamenti climatici e, di particolare rilevanza per il progetto in corso, come supporto computazionale alle scienze della vita. Al consorzio per il progetto E4C partecipano anche i centri di Barcellona (BSC), Juelich (JSC) e Stoccolma (KTH) e, in ogni caso, tutta la rete dei supercalcolatori europei si è detta disponibile a mettere a disposizione le proprie macchine per accelerare le simulazioni previste dal progetto.

Nel fronteggiare un evento critico come l'attuale pandemia, il tempo ha un ruolo strategico. Che cosa s'intende con 'urgent computing'?

[Carlo Cavazzoni, Cineca] Diciamo che, almeno fino a oggi, la definizione di *urgent computing* non era chiara ai più, in quanto legata a un concetto del tutto astratto. Da questo punto di vista, l'attuale crisi legata a COVID-19 è il primo caso reale che si presenta di *urgent computing*, ovvero la possibilità di allocare in brevissimo tempo tutta la potenza di calcolo disponibile destinandola a una singola applicazione. In questo caso, l'applicazione consiste nell'elaborazione dei modelli che consentano di individuare i farmaci adatti a fronteggiare l'emergenza COVID-19. L'obiettivo, che giustifica la focalizzazione totale delle risorse, è contribuire a mitigare gli effetti di una catastrofe che mette a rischio la vita di milioni di persone e che richiede una risposta immediata e urgente.

Come è nata la collaborazione del Cineca con l'azienda farmaceutica Dompé? E Come funziona la piattaforma Exscalate?

[Carlo Cavazzoni, Cineca] La collaborazione tra Cineca e Dompé è nata ben quindici anni fa, quando si è cominciato a capire che grazie alla potenza dei supercalcolatori era possibile costruire modelli digitali sufficientemente accurati delle proteine e delle molecole dei farmaci. La potenza sarebbe arrivata (come è stato) con l'evoluzione dei microprocessori, mentre mancava ancora il software o, meglio, quello che c'era non era sufficientemente accurato e veloce. Così cominciammo a sviluppare un *software* (LiGen) per la progettazione dei farmaci al calcolatore, che fosse disegnato secondo i principi del calcolo parallelo, in modo da sfruttare la potenza dei moderni supercalcolatori. Un'altra caratteristica del disegno del software che

» INTERVISTA

abbiamo sviluppato con Dompé è quella di essere modulare ed estensibile se integrato con altri software che possano assolvere ad altre funzioni del flusso di lavoro necessario a completare il modello digitale della proteina e del farmaco. Exscalate è appunto un insieme di flussi di lavoro codificati che includono LiGen, GEODOCK (un componente sviluppato dal Politecnico di Milano), Gromacs (un codice di dinamica molecolare sviluppato dal KTH di Stoccolma) e altri componenti necessari alla gestione informatica dei flussi di lavoro. In questo modo Exscalate rende automatica tutta una serie di passaggi garantendo rapidità e riproducibilità dei risultati, ovvero quello che è necessario in casi come quello legato a COVID-19.

Quando sono attesi i primi risultati? E quando potranno essere applicati realmente in ambito farmaceutico?

[Carlo Cavazzoni, Cineca] Dal punto di vista dei modelli numerici di potenziali farmaci che possano rallentare l'azione del virus COVID-19, abbiamo già trovato una serie di possibili candidati tra i farmaci e le molecole già impiegate sull'uomo. Questo non significa che queste molecole possano già essere impiegate in ambito clinico, occorre prima che siano testate in laboratorio per confermare in-vitro quello che è stato ottenuto "in-silico", con il calcolatore, ovvero l'azione del farmaco sulla proteina. In ogni caso, stimiamo che questa fase urgente di verifica di tutte le possibili molecole note già impiegate sull'uomo possa durare un paio di mesi. Contiamo, in questo tempo, di poter produrre circa un centinaio di buoni candidati da testare in laboratorio e poi sull'uomo. Se questo non bastasse, la piattaforma è pronta anche per la progettazione di nuove molecole e farmaci ancora sconosciuti. In questo caso però i tempi per l'applicazione all'uomo sarebbero più lunghi, perché si renderebbe necessario verificare che le nuove molecole siano sicure per l'uomo.

Dr. Dell'Agnello, il CNAF di Bologna è il principale centro di calcolo dell'INFN: un'infrastruttura dotata di risorse di calcolo d'avanguardia, necessarie all'elaborazione dei milioni di dati prodotti dagli esperimenti di fisica delle alte energie e di fisica astroparticellare. Con quali competenze specifiche l'INFN partecipa al consorzio?

[Luca dell'Agnello, INFN-CNAF] Le competenze del CNAF sono in una qualche misura complementari a quelle del CINECA. In particolare, oltre a fornire ai nostri esperimenti una potenza di calcolo quantificabile in centinaia di migliaia di processi al giorno, abbiamo una capacità di stoccaggio dei dati di decine di PetaByte sia on-line (su disco) che sulle librerie a nastro per l'archiviazione a lungo termine. Uno dei nostri punti di forza è proprio il *Data Management*, frutto dell'esperienza decennale maturata sviluppando e fornendo soluzioni d'avanguardia agli esperimenti all'acceleratore LHC del CERN e ad oltre trenta altri esperimenti ai quali l'INFN partecipa. Oggi, per la collaborazione Exscalate, stiamo mettendo in campo, oltre alla potenza

» INTERVISTA

di calcolo, un sistema di memorizzazione basato su tecnologia *Object Storage* in grado di accogliere i dati delle simulazioni prodotte al CINECA e renderle poi fruibili a tutto il Consorzio: questo tipo di competenze è proprio quello sviluppato nel corso degli anni trattando i dati degli esperimenti di fisica delle alte energie e di fisica astroparticellare. Infatti, la quantità di dati prodotta dagli esperimenti, che viene scritta e letta quotidianamente, può raggiungere oggi anche il centinaio di TeraByte. Il CNAF, inoltre, è uno dei centri di calcolo più performanti a livello di velocità di accesso ai dati. In termini di quantità di dati trattati, prevediamo di superare i 100 PetaByte nel 2025.

Sempre nell'ambito di trattamento dei dati sperimentali, la sfida che il CNAF sta affrontando attualmente è la preparazione per *High-Luminosity LHC* (il quarto Run di LHC): dal 2027, verrà prodotta una quantità di dati che supererà l'attuale di almeno 10 volte. Questo ci pone di fronte a un'importante sfida tecnologica, soprattutto per quanto riguarda l'accesso e l'elaborazione dei dati.

Come è nata l'attuale concentrazione di attività, di tecnologie e di eccellenza nel supercalcolo, a Bologna?

[Luca dell'Agnello, INFN-CNAF] Bologna è sempre stato uno dei poli di eccellenza dell'informatica in Italia: ne sono un esempio la storia del CINECA e a quella del CNAF. In particolare, mentre il CINECA si è specializzato nel calcolo parallelo, negli anni '90 del secolo scorso il CNAF ha svolto la funzione di incubatore della rete italiana della ricerca (il GARR) e, al volgere del millennio, ha affrontato la sfida del calcolo per gli esperimenti all'acceleratore LHC del CERN. All'epoca, la sfida ci poneva di fronte a un problema arduo, sia per la potenza di calcolo richiesta sia per la gestione dei dati, e non solo per la loro mole ingente, ma anche per l'efficienza della loro movimentazione. Per questo motivo, a partire dal 2000, il CNAF ha partecipato a numerosi progetti europei e nazionali per lo sviluppo di un'infrastruttura di calcolo per gli esperimenti di fisica, che fosse basata sul calcolo distribuito invece che sul calcolo parallelo, contribuendo così a progettare e a implementare servizi all'avanguardia. A partire dal 2003, al CNAF è stato realizzato il principale data center dell'INFN, il Tier-1, originariamente destinato a supporto di calcolo dei quattro principali esperimenti a LHC e che attualmente fornisce servizio a circa 40 esperimenti. Un ulteriore fattore di rilievo è l'integrazione del data center del CNAF nell'infrastruttura di calcolo distribuito estesa a livello mondiale, in grado di soddisfare oltre alle esigenze della fisica delle alte energie e astroparticellare, quelle di altre comunità scientifiche.

Ha preso vita a Bologna una vera e propria cittadella del supercalcolo.

[Luca dell'Agnello, INFN-CNAF] Sì. A Bologna sono attualmente presenti i due data center pubblici più importanti ma con caratteristiche complementari: il CINECA specializzato nel supercalcolo HPC (*High*

» INTERVISTA

Performance Computing), il CNAF nel calcolo HTC (*High Throughput Computing*) e nel *data management*. Un passo importante verso una più stretta collaborazione tra i due centri è avvenuto con la scelta di Bologna, come sede del nuovo data center dell'ECMWF (*European Center for Medium-Range Forecasts*), il Centro Europeo di Previsioni Meteo a Medio Termine che fornisce previsioni meteorologiche agli Stati membri. È stato il primo passo per la nascita, sotto il patrocinio della Regione Emilia-Romagna, del progetto Tecnopolo che vedrà sorgere presso l'ex Manifattura Tabacchi di Bologna, oltre al data center dell'ECMWF, anche il nuovo data center del CNAF e del CINECA. In particolare, sarà installato Leonardo, una delle tre nuove macchine pre-Exascale finanziate dalla Commissione Europea, e saranno trasferite le risorse di calcolo e di stoccaggio dati del CNAF, destinate a crescere in modo sostanziale nella seconda metà del decennio. Questo favorirà una più stretta collaborazione tra CNAF e CINECA realizzando uno dei più grandi centri di calcolo a livello mondiale.

Quali sono i vantaggi attesi dall'integrazione di competenze biomediche e di supercalcolo? Perché si è resa necessaria la costituzione di un consorzio di 18 istituzioni e centri ricerca, di 7 paesi europei, afferenti a diversi campi di ricerca?

[Luca dell'Agnello, INFN-CNAF] Data l'esperienza ventennale nel campo del calcolo scientifico d'avanguardia, l'INFN è in condizione di mettere il proprio *know-how* a disposizione di altre discipline scientifiche, come dimostrato dalle numerose collaborazioni già in essere, attraverso progetti nazionali e internazionali, con comunità bioinformatiche, mediche e farmaceutiche. Anche per questo progetto, la sfida che abbiamo davanti, e che possiamo definire vitale, richiede di mettere in campo, da una parte, le eccellenze europee nel settore biomedico, dall'altra gli specialisti della bioinformatica per i programmi di simulazione per la ricerca di farmaci in grado di neutralizzare il Covid-19. Altrettanto fondamentale, però, è l'apporto delle grandi infrastrutture che possono fornire la potenza di calcolo indispensabile per questo tipo di ricerca: solo con l'uso dei supercalcolatori è possibile abbreviare i tempi per lo sviluppo di farmaci. In sintesi, l'interdisciplinarietà, sempre feconda, è una delle chiavi per il successo in un campo d'avanguardia come questo. Come CNAF, il nostro compito sarà quello di supportare i ricercatori non solo fornendo tutti gli strumenti informatici di cui necessiteranno ma anche coadiuvandoli facilitandone l'uso. ■

* Il consorzio Exscalate4CoV, guidato da Dompé farmaceutici, è costituito da 18 istituzioni e centri di ricerca di 7 Paesi europei: Politecnico di Milano (Dipartimento di Elettronica, Informazione e Bioingegneria), Consorzio Interuniversitario CINECA (Supercomputing Innovation and Applications), Università degli Studi di Milano (Dipartimento di Scienze Farmaceutiche), Katholieke Universiteit Leuven, International Institute Of Molecular And Cell Biology In Warsaw (LIMCB), Electra Italian Crystallographic Association, Fraunhofer Institute for Molecular Biology and Applied Ecology, Bsc Supercomputing Centre, Forschungszentrum Jülich, Università Federico II di Napoli, Università degli Studi di Cagliari, SIB Swiss Institute of Bioinformatics, KTH Royal Institute of Technology (Department of Applied Physics), Associazione BigData, Istituto Nazionale Di Fisica Nucleare (INFN), l'Istituto nazionale per le malattie infettive Lazzaro Spallanzani e Chelonia Applied Science.