

High Performance Computing (HPC):

note introduttive

di Giuseppe F. Italiano

1. Introduzione

Viviamo in un mondo sempre più digitale. In particolare, in questo momento stiamo attraversando una profonda rivoluzione digitale: oggi si generano decine di zettabyte (decine di migliaia di miliardi di gigabyte) all'anno, con un tempo di raddoppio (il tempo necessario per generare una quantità di dati pari a quella prodotta nell'intera storia fino a quel momento) di circa un anno. Si prevede che entro 3-4 anni, circa 150 miliardi di nuovi dispositivi "intelligenti" saranno connessi a Internet e di conseguenza i dati generati annualmente aumenteranno di diversi yottabyte, un numero addirittura superiore alla costante di Avogadro (che esprime il numero di particelle – atomi, molecole o ioni – contenute in una mole),

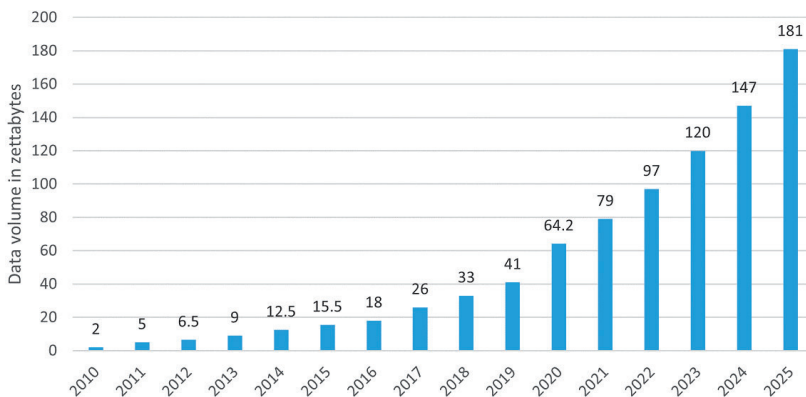


Fig. 1 – Volume di dati per anno (in zettabytes)

mentre il tempo di raddoppio scenderà a poche ore. Come si vede dalla Figura 1 (Fonte Statista 2022), il volume di dati creati, elaborati, copiati e consumati nel mondo in un anno sta infatti crescendo con una velocità esponenziale.

Con il crescere della disponibilità di dati, sta crescendo in tutto il mondo anche il mercato dei data analytics, che assume una particolare rilevanza anche in Italia. Come evidenziato nella Figura 2 (Fonte Osservatorio Digitale Transformation Academy), a parte una lieve flessione nel 2020 dovuta alla pandemia da Covid-19, negli ultimi 5 anni in Italia il mercato degli analytics ha avuto tassi di crescita annuali a doppia cifra.

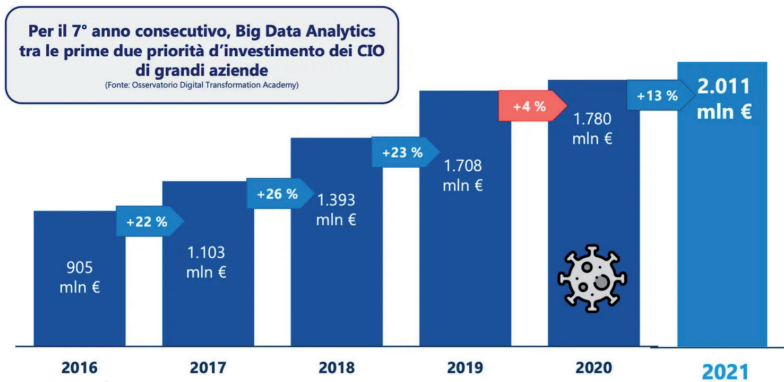


Fig. 2 – Il mercato degli analytics in Italia

In questo scenario l'intelligenza artificiale, e in particolare la *machine learning*, con i suoi progressi incredibili degli ultimi decenni, sembra essere l'unico strumento possibile per analizzare ed estrarre valore da grandi volumi di dati. In particolare, la *machine learning* è ormai essenziale in tutte le fasi dell'analisi di big data, che vanno dall'estrarre pattern dai dati (informazione), all'individuare correlazioni tra tali pattern (conoscenza), all'utilizzare tale conoscenza per costruire una

rappresentazione matematica (un modello) coerente con i dati e infine alla realizzazione di algoritmi predittivi coerenti con tali modelli.

Elaborare grandi quantità di dati con algoritmi di *machine learning* ha comportato una crescente richiesta di potenza di calcolo. Si stima infatti che dal 2012 la quantità di calcolo utilizzata nelle fasi di training da algoritmi di *machine learning* sia aumentata in modo esponenziale, con un periodo di raddoppio (*doubling time*) pari a 3,4 mesi (ovvero la quantità di calcolo utilizzata nella fase di training raddoppia all'incirca ogni 3,4 mesi). Tanto per fare un confronto, la Legge di Moore prevede un periodo di raddoppio pari a 2 anni. Questo trend relativo alla crescente domanda di capacità di calcolo viene illustrato in dettaglio in Figura 3 (Fonte <https://openai.com/blog/ai-and-compute/#addendum>). Per fissare le idee, dal 2012 la metrica relativa alla quantità di calcolo utilizzata

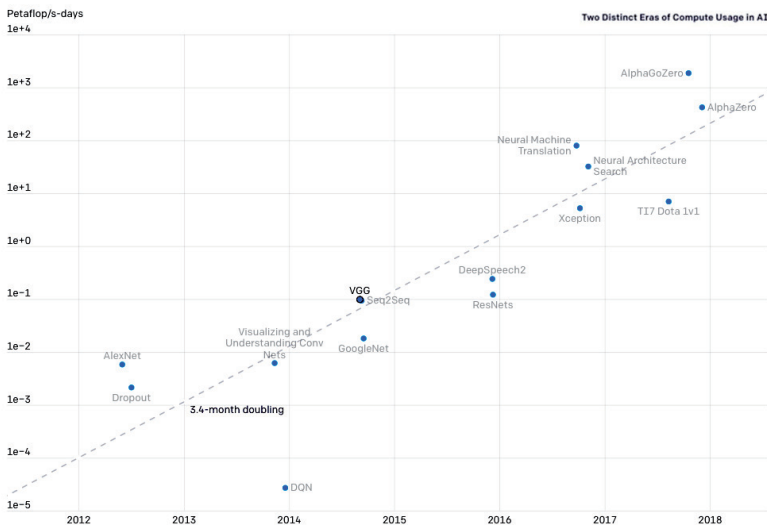


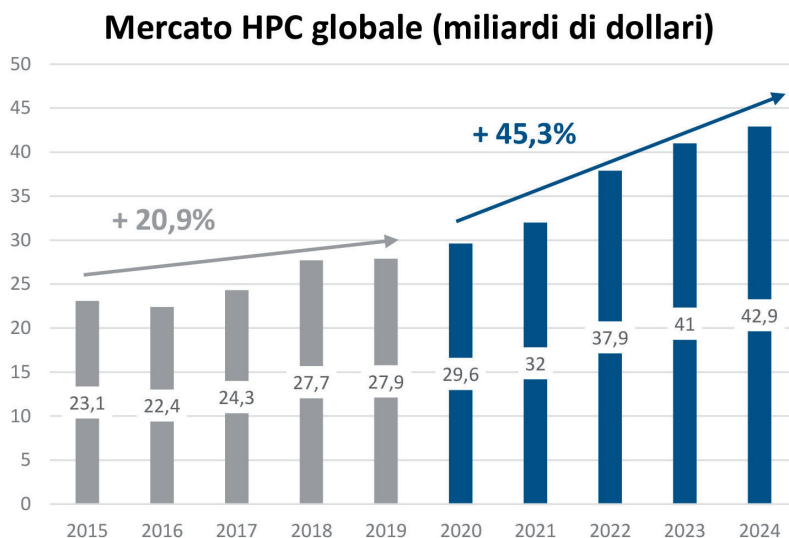
Fig. 3 – Capacità di calcolo utilizzata da algoritmi di *machine learning* nella fase di training

è cresciuta di oltre 300 mila volte (il raddoppio di 2 anni previsto dalla Legge di Moore produrrebbe soltanto una crescita di un fattore 7!). Questi miglioramenti nella capacità di elaborazione sono stati una componente chiave del progresso degli algoritmi di *machine learning*. Finché questo trend continua, è importante essere pronti a prevedere sistemi di calcolo in grado di andare molto al di là delle capacità di calcolo odierne.

2. High Performance Computing (HPC)

Con High Performance Computing (calcolo a elevate prestazioni) ci si riferisce alle tecnologie utilizzate (oggi in gran parte computer cluster) per creare sistemi di elaborazione in grado di fornire prestazioni molto elevate (attualmente dell'ordine dei petaflops, ricorrendo tipicamente a calcolo parallelo). L'interesse per HPC è iniziato negli anni Sessanta, con le prime applicazioni, soprattutto relative a simulazioni, nate nelle università e negli enti governativi statunitensi. Negli anni successivi, l'evoluzione delle architetture hardware e degli strumenti software ha consentito di sfruttare al meglio le potenzialità dei sistemi HPC per una molteplicità di applicazioni. Negli ultimi anni, infatti, sono emerse moltissime nuove possibilità di applicazione di HPC in contesti aziendali, molto diversi da contesti originari per cui era stato concepito HPC. Tra le principali applicazioni odierne di HPC citiamo la sicurezza cyber, la biometria, l'ottimizzazione, la gestione del rischio, il miglioramento dei processi, la previsione comportamentale e il miglioramento di business model. Anche grazie alle innumerevoli applicazioni in contesti aziendali, il mercato HPC globale è in forte crescita e si prevedono ulteriori sviluppi, come riportato in Figura 4 (Fonte Statista).

A mero titolo di esempio di applicazioni HPC in contesti aziendali, citiamo PayPal che utilizza un cluster HPC per identificare frodi e pattern di comportamenti sospetti in tem-



Fonte: Statista

Fig. 4 – Il mercato HPC globale (in miliardi di dollari)

po reale; per creare nuovi modelli anti-frode in modo rapido e con minimo effort; per estrapolare insight in tempi rapidi dai dati e dagli eventi che vengono registrati continuamente sui sistemi aziendali. PayPal realizza questi obiettivi riuscendo a correlare dati provenienti da molte sorgenti eterogenee e ottenendo così informazioni utili a pianificare interventi di business tempestivi. In particolare, la velocità di calcolo del cluster HPC consente di elaborare e correlare dati provenienti da migliaia di sorgenti diverse, relativi a circa 3 milioni di eventi/secondo, così da riuscire ad avere insight in tempo reale. Grazie a questo cluster HPC, PayPal è in grado di analizzare application log, dati operativi, ambientali ed eventi registrati su social media, con flussi di 25 Tb/ora, provenienti da migliaia di server, inclusi i trend su social media e interazioni clienti. Tutto questo consente di identificare pattern e anomalie: grazie all'analisi dei dati in tempo reale HPC permette

all'azienda di intervenire immediatamente, prima di eventuali esperienze negative da parte degli utenti.

Cambiando segmento di business, ENI SpA utilizza HPC in attività upstream, sia per migliorare la ricerca di idrocarburi (tramite la modellazione dei giacimenti in fase di produzione), sia per migliorare le capacità di predizione e di simulazione. Per raggiungere questi obiettivi, ENI ha due sistemi HPC, con capacità di calcolo di picco combinata di 52 petaflop/s, in cui combina simulazione numerica con tecniche di *machine learning*, e in particolare *deep learning*. La capacità di calcolo dei sistemi HPC ha permesso a ENI di ridurre i tempi necessari a ricostruire i modelli del sottosuolo da alcuni mesi a pochi giorni o addirittura a ore. Inoltre, con l'introduzione dei sistemi HPC, sono stati dimezzati i tempi di completamento della prima produzione, passando da 9 anni a 4,5 anni.

Come riportato nella Figura 5 (Fonte www.top500.org), con dati aggiornati a giugno 2022, i paesi leader mondiali

	Countries	Count	System Share (%)	Rmax (GFlops)	Rpeak (GFlops)	Cores
1	China	173	34.6	530,240,332	1,158,770,698	29,413,676
2	United States	128	25.6	2,085,045,140	3,150,397,946	27,715,304
3	Japan	33	6.6	626,505,930	817,352,794	11,984,068
4	Germany	31	6.2	200,537,160	306,053,664	3,896,660
5	France	22	4.4	168,660,070	242,483,550	3,874,520
6	Canada	14	2.8	47,804,800	80,390,249	1,076,384
7	United Kingdom	12	2.4	57,018,004	78,628,972	1,779,888
8	Russia	7	1.4	73,715,000	101,737,460	741,328
9	Italy	6	1.2	78,529,000	114,511,528	1,447,536
10	Netherlands	6	1.2	27,627,110	44,565,960	398,928
11	Brazil	6	1.2	24,113,500	39,212,939	460,736
12	Saudi Arabia	6	1.2	55,253,040	98,982,254	1,798,260
13	South Korea	6	1.2	83,677,000	118,233,580	1,460,292
14	Poland	5	1	14,321,830	22,271,790	234,960
15	Australia	5	1	17,196,950	27,300,835	408,768
16	Sweden	5	1	14,241,500	22,519,838	302,784
17	Switzerland	4	0.8	28,563,790	38,600,165	650,260
18	Finland	4	0.8	165,295,040	231,727,305	1,514,432
19	Singapore	3	0.6	8,570,830	14,425,080	222,208
20	India	3	0.6	10,953,340	12,082,026	244,488

Fig. 5 – Installazioni HPC nel mondo (giugno 2022)

nell’HPC sono la Cina e gli Stati Uniti, rispettivamente con 173 e 128 installazioni. L’Italia occupa attualmente il nono posto nella classifica mondiale con 6 installazioni.

In Figura 6 (Fonte www.top500.org) si riportano i 5 sistemi HPC attualmente più potenti al mondo, sempre con dati aggiornati a giugno 2022. Da maggio 2022 abbiamo un nuovo numero 1, il sistema HPC Frontier installato presso l’Oak Ridge National Laboratory (ORNL), Tennessee, USA. Frontier è il primo sistema HPC exascale al mondo e riporta la prima posizione negli Stati Uniti dopo che è stata detenuta per 2 anni dal sistema Fugaku al RIKEN Center for Computational Science (R-CCS) a Kobe, in Giappone.

Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)
1	Frontier - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,730,112	1,102.00	1,685.65	21,100
2	Supercomputer Fugaku - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899
3	LUMI - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	1,110,144	151.90	214.35	2,942
4	Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM DOE/SC/Oak Ridge National Laboratory United States	2,414,592	148.60	200.79	10,096
5	Sierra - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA / Mellanox DOE/NNSA/LLNL United States	1,572,480	94.64	125.71	7,438

Fig. 6 – I sistemi HPC più potenti al mondo (giugno 2022)

A giugno 2022 c'è stata una new entry anche al numero 3, il sistema HPC LUMI installato presso EuroHPC/CSC in Finlandia, che è attualmente il più grande sistema in Europa. La classifica riporta anche un terzo nuovo arrivato nella Top 10: al numero 10 appare infatti il sistema Adastra del GENCI-CINES in Francia, a testimonianza dello sforzo che l'Unione Europea sta facendo in ambito HPC. Sempre nell'ultima classifica disponibile (giugno 2022), l'Italia aveva tre sistemi HPC nella Top 100 mondiale: HPC5 di ENI SpA in dodicesima posizione, Marconi-100 del CINECA in ventunesima posizione e HPC4 di ENI SpA in quarantaquattresima posizione.

3. HPC nell'Unione Europea

HPC rappresenta indubbiamente un asset di garanzia per l'autonomia tecnologica europea. Secondo le stime di European House Ambrosetti (2018), l'Unione Europea consuma il 25% delle risorse di supercalcolo mondiali, fornendone però soltanto il 5%. Allo scopo di ridurre livello di dipendenza dalle tecnologie extra-UE, la Commissione Europea ha sviluppato una strategia HPC che si sta muovendo principalmente lungo tre direzioni:

- creare un'industria europea capace di sviluppare componenti e tecnologie HPC “made in EU”;
- sviluppare una visione di lungo periodo con l'obiettivo di diffondere le tecnologie HPC all'interno del tessuto economico europeo;
- rendere l'Unione Europea autonoma dal punto di vista della potenza di calcolo installata, così da aumentare sicurezza, privacy e gestione delle informazioni generate da sistemi HPC.

Per ottenere risultati concreti in queste tre direzioni, negli anni recenti l'Unione Europea ha finanziato vari programmi, tra cui citiamo in particolare:

- EPI (European Processor Initiative), con l'obiettivo di sviluppare le componenti hardware dei sistemi HPC all'interno della Comunità Europea;
- PRACE (Partnership for Advanced Computing in Europe), con l'obiettivo di favorire la diffusione della ricerca e lo sviluppo di soluzioni ad alto impatto scientifico che possano aumentare la competitività dell'Europa, generando benefici per l'umanità;
- EuroHPC Joint Undertaking, con l'obiettivo di rendere l'Europa autonoma dal punto di vista della potenza computazionale installata e sviluppare la prossima generazione di supercomputer exascale;
- SHAPE (SME HPC Adoption Programme in Europe), con l'obiettivo di facilitare l'accesso alla tecnologia HPC alle piccole e medie imprese.

4. HPC in Italia

L'Italia è il terzo Paese dell'Unione Europea per potenza di calcolo, dopo Germania e Francia, con il 18,4% della potenza totale europea. Negli ultimi anni la potenza di calcolo italiana è aumentata del 354%, passando da 17.308 Tflop/s a oltre 78.529 Tflop/s. Rispetto agli altri contesti di riferimento (Unione Europea e Mondo), l'Italia si distingue per il maggiore utilizzo di HPC in ambito industriale.

Ad oggi, come riportato precedentemente, i maggiori attori del mondo del supercalcolo italiano, censiti da Top 500, sono ENI S.p.A. e CINECA. Tra gli altri attori è importante citare anche altre 3 organizzazioni:

- Dallara, che opera nel settore dell'industria automobilistica da competizione, e che utilizza HPC per la progettazione e la realizzazione di componenti High Performance;
- Dompé, che opera nel settore biopharma, e che utilizza HPC soprattutto per il Computer Aided Drug Design

(CADD) attraverso una piattaforma dedicate di drug discovery;

- European Centre for Medium-Range Weather Forecasts, con sede a Bologna, che usa HPC per fornire previsioni meteo accurate in Europa.

Nel futuro immediato si prevedono due grosse nuove installazioni di HPC in Italia. La prima è davinci-1 di Leonardo SpA, un sistema HPC basato su tecnologia Atos e acceleratori GPU Nvidia. Una volta ultimato, con i suoi 248.000 Tflop/s davinci-1 diventerà uno dei 5 sistemi HPC più potenti al mondo e farà aumentare la potenza di supercalcolo italiana di 4,2 volte, facendo scalare all'Italia la classifica mondiale di capacità computazionale, portandola al quarto posto. Si prevede che 50% della potenza di calcolo di davinci-1 sarà resa disponibile ai ricercatori e alle imprese italiane così che anche loro possano sfruttare le potenzialità offerte dalla tecnologia HPC. I campi di applicazione in cui potrà essere utilizzato davinci-1 sono molteplici: drug discovery, space exploration and research, weather modeling e individuazione delle strategie per la mitigazione di eventi naturali e di situazioni di pandemia.

La seconda installazione HPC verrà effettuata nell'ambito del progetto EuroHPC JU (European High Performance Computing Joint Undertaking), citato in precedenza. In particolare, l'Italia ospiterà Leonardo, uno dei supercomputer "pre exascale" (precursore della tecnologia "exascale"), che sarà collocato presso il Tecnopolo di Bologna (Casalecchio di Reno) e sarà gestito dal CINECA, e che è stato finanziato con 240 milioni di euro dalla Commissione Europea e dal Governo italiano. Leonardo avrà molteplici applicazioni, tra cui la riduzione e la gestione rischi naturali (terremoti, tsunami, eruzioni vulcaniche, pandemie, etc.), la medicina personalizzata e predittiva, le previsioni meteo, la sicurezza cyber, e l'intelligenza artificiale.

5. HPC e Quantum Computing

Il *Quantum Computing* è un modello di calcolo basato sulla teoria quantistica. Uno degli aspetti più interessanti dal punto di vista computazionale è il bit quantistico (Qubit), l'unità di informazione quantistica che può esistere contemporaneamente in due stati (polarizzazione orizzontale e verticale), grazie al principio di sovrapposizione della fisica quantistica. Il potere computazionale del *Quantum Computing* non si basa quindi su velocità di clock più elevate, ma sul fatto che può gestire dati esponenzialmente più grandi (es. fattorizzazione). Dati i limiti attuali del *Quantum Computing* e l'evoluzione di HPC, nel breve termine si possono prevedere sistemi ibridi che utilizzano tecnologie quantistiche e classiche sulla stessa piattaforma: effettuando elaborazioni general purpose ad alta velocità (con HPC) combinate con ultra-performance specifiche per i casi d'uso (con *Quantum Computing*).

In particolare, il programma EuroHPC JU (European High Performance Computing Joint Undertaking) ha finanziato il progetto HPCQS (High Performance Computer and Quantum Simulator hybrid), con l'obiettivo di integrare tecnologie HPC quantistiche e classiche (integrando simulatori quantistici nei supercomputer europei già esistenti). Questo rappresenterà un incubatore per Quantum-HPC hybrid computing unico al mondo: potrà liberare nuovo potenziale innovativo e preparare Europa per l'era post exascale.

6. HPC e Cloud

Nell'immediato futuro è anche importante avviare un'ulteriore riflessione sulle possibili interazioni tra HPC e cloud computing. Negli ultimi anni, infatti, i principali fornitori di cloud hanno investito in reti globali di sistemi massive-scale che hanno il potenziale di essere molto competitivi con gli attuali sistemi HPC. Spinti dalle enormi richieste di calcolo

degli algoritmi di intelligenza artificiale a cui abbiamo precedentemente accennato, gli odierni sistemi cloud sono costruiti sempre più utilizzando chip custom e semiconduttori personalizzati. Questo rischia di avere ripercussioni negative su HPC, perché ha l'effetto di ridurre il leverage finanziario dei vendor tradizionali di hardware e di CPU, che hanno da sempre avuto un ruolo centrale nello sviluppo di sistemi HPC. Nell'immediato futuro saranno quindi gli attuali fornitori di cloud ad assicurarsi i servizi e le applicazioni tipiche dei sistemi HPC? Difficile dirlo, anche se già da oggi le architetture e i sistemi cloud stanno assumendo posizioni di assoluta leadership in molte applicazioni, come ad esempio il gaming e la visione artificiale, e di conseguenza stanno ridefinendo anche il modo in cui pensiamo alla natura stessa del calcolo ad alte prestazioni. Per questi motivi costruire la prossima generazione di sistemi HPC richiederà molto probabilmente un ripensamento profondo degli approcci tradizionali, e una maggiore attenzione alle caratteristiche di successo delle architetture cloud, come ad esempio l'adozione di configurazioni hardware personalizzate e una maggiore prototipazione su larga scala. Vi è inoltre una ulteriore importante considerazione da tenere presente: fino ad oggi, i sistemi HPC hanno da sempre necessitato di grossi investimenti finanziari da parte dei governi e dal mondo della ricerca, mentre al contrario i sistemi cloud sono in grado di generare grosse revenue. È questo ancora un modello sostenibile per i sistemi HPC? Anche questo aspetto sembra suggerire la necessità di un rafforzamento delle collaborazioni tra HPC e gli attori che hanno una posizione dominante nell'attuale ecosistema informatico, inclusi i fornitori di cloud computing.