



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

AI-based Predictive Modelling and Optimization for Sustainable High-Performance Computing

Progetto di ricercare

I sistemi di High-Performance Computing (HPC) svolgono un ruolo fondamentale nell'assistere i progressi scientifici in diversi ambiti, dagli studi epidemiologici alla scienza del clima. Con la crescita esponenziale delle richieste computazionali, è essenziale affrontare l'impatto ambientale legato all'impronta di carbonio e al consumo energetico dei sistemi. Una soluzione efficace consiste nello sviluppare e integrare tecniche basate sull'intelligenza artificiale per abilitare strategie automatizzate sulla gestione del workload del sistema. Modelli avanzati di machine learning combinati con strumenti di natural language processing, insieme ad algoritmi di ottimizzazione, possono essere impiegati per prevedere le caratteristiche del workload del sistema e adattare dinamicamente la gestione del workload in tempo reale. Tale soluzione ha il potenziale per ottimizzare contemporaneamente il throughput del sistema e l'impatto ambientale. L'obiettivo di questo progetto è di offrire soluzioni per rendere più ecologiche ed efficienti le infrastrutture HPC, per un futuro digitale più verde e sostenibile.

Piano di attività

Gli obiettivi del progetto saranno raggiunti tramite lo sviluppo di algoritmi predittivi basati su ML per il workload di un sistema HPC (composto da jobs sottomessi dagli utenti) e integrandoli nelle strategie di gestione del workload. Gli algoritmi sfrutteranno pipeline RAG end-to-end con LLMs per prevedere le caratteristiche dei jobs (come durata, consumo energetico, impronta di carbonio), basandosi sulle informazioni disponibili al momento della sottomissione del job (come job script e codice sorgente) e sui log raccolti durante l'esecuzione. Queste informazioni verranno poi utilizzate per elaborare strategie di gestione del workload basate su algoritmi di ottimizzazione a vincoli, al fine di ottimizzare contemporaneamente il throughput del sistema e l'impatto ambientale (come l'impronta di carbonio e il consumo energetico).