

Progetto per Assegno di Ricerca

"Infrastrutture CloudHPC per il supporto alla ricerca"

Scenario

L'obiettivo del progetto è quello di realizzare un'infrastruttura per la sperimentazione sul campo di diverse soluzioni architetturali volte a supportare le esigenze di calcolo e memorizzazione di dati, partendo dalla ricognizione delle migliori pratiche adottate da altri Atenei paragonabili all'Alma Mater, che ha individuato come fattori comuni: l'adozione ubiqua di strumenti open source per l'allocazione delle risorse e il tracciamento delle attività; modalità di accesso offerta all'utenza con un alto grado di automazione; in alcuni casi, tariffazione dell'uso delle risorse.

L'architettura di base sarà quella di un cloud ibrido, gestito da uno strato software composto di due livelli interoperanti: uno permetterà la gestione distribuita di risorse di rete, calcolo e storage secondo il modello cloud, l'altro gestirà l'allocazione di task di calcolo sui nodi esposti dal primo. Il risultato complessivo dovrà essere un'infrastruttura che integri ogni variante di risorsa disponibile. Si prevede di acquisire un nucleo locale di dimensioni adeguate alle richieste correnti, strutturato per essere facilmente espandibile in futuro, compensando i picchi di domanda transitori con l'utilizzo di risorse su cloud pubblico di partner convenzionati.

Requisiti

Lo svolgimento del progetto sarà incentrato sulla realizzazione del suddetto strato software, affinando l'analisi dei requisiti, e interagendo con l'utenza per verificare l'efficacia dei risultati, e prevede l'applicazione di competenze:

- di integrazione e sviluppo di sistemi per l'allocazione di job su architetture di calcolo distribuite
- di sicurezza informatica, dimostrabile anche con la partecipazione a competizioni internazionali
- di sviluppo software in ambienti eterogenei, anche web e mobile, per la realizzazione dell'interfaccia utente
- di interazione con l'utenza universitaria per la raccolta dei requisiti e l'affiancamento nelle fasi di test del prototipo realizzato

Piano delle attività

1. Analisi dei requisiti dei gruppi di ricerca, in termini di tipologia di task di calcolo, di vincoli temporali per l'esecuzione dei task , di risorse necessarie per il calcolo e per la memorizzazione dei dati(quantità e modalità di accesso);
2. Studio delle soluzioni disponibili a mercato in termini di nodi di calcolo ad alte prestazioni, sistemi di elaborazione a supporto della virtualizzazione, sistemi di storage gerarchici per applicazioni HPC, Cloud, e archiviazione a lungo termine di dataset dell'ordine dei PetaByte;
3. Studio dei sistemi per l'allocazione flessibile e dinamica delle risorse, e individuazione delle piattaforme software per la realizzazione dei due livelli di gestione interoperanti (allocazione dei sistemi virtualizzati, distribuzione dei task);
4. Implementazione e test di un prototipo gestito da sistemisti esperti per la validazione dei modelli di allocazione delle risorse;
5. Analisi e progetto dei connettori verso
 - a. differenti architetture hardware per poter federare datacenter distribuiti
 - b. sistemi cloud pubblici per il trasferimento di carichi di lavoro all'esterno
 - c. sistemi di gestione delle identità di Ateneo e federate
6. Analisi e progetto delle policy per l'utilizzo, includendo
 - a. modalità di amministrazione (centralizzata, supportata, delegata)
 - b. modalità di accounting, tariffazione, incentivazione a contribuire risorse
 - c. sviluppo della documentazione e del supporto all'utenza
 - d. schemi di controllo dell'accesso
7. Progetto e realizzazione di interfacce utente che massimizzino l'indipendenza nell'uso dell'infrastruttura, sia sotto forma di API per poterne integrare la gestione in processi automatizzati, sia sotto forma di viste web per consentirne l'uso a utenti non specializzati.