

## Prefazione

La seguente trattazione presenta un progetto atto allo studio dell'efficienza computazionale di cluster a basso consumo energetico adoperati per l'analisi biofisica. In particolare, la ricerca mira a dimostrare come l'utilizzo di macchine a minor dispendio energetico possano essere pi convenienti e potenzialmente pi potenti rispetto alle odierne macchine sfruttate nel ramo della ricerca biomedica e sanitaria.

Il processo attraverso il quale stato possibile strutturare una tale ricerca avvenuto concentrando gli interessi verso uno tra i sistemi pi moderni di ricerca delle mutazioni genetiche causanti varie tipologie di tumori: il sistema GATK-LOD<sub>n</sub>. L'implementazione di una componente di questo metodo in un innovativa versione della nota utility Make e, quindi, la creazione di un nuovo eseguibile, ha permesso una gestione pi libera dei singoli passaggi del suddetto sistema.

Questo nuovo strumento, denominato Snakemake, in grado di organizzare i diversi compiti del metodo biofisico su diverse macchine, conservando il corpo unico del programma. Proprio questa caratteristica stata sfruttata per verificare la fruibilit di insiemi di computer che collaborano come un solo apparato: i cluster.

La formazione di gruppi di computer nasce dalla necessit dei moderni organi di ricerca. In questi tempi gli studi nel campo biomedico prevedono, generalmente, la collaborazione di professionisti in statistica dotati di computer ad alta potenza, che hanno il compito di fornire i risultati proficuamente e rapidamente. Lo sviluppo e il progresso nei vari settori ha necessitato, e necessita tuttora, di un contemporaneo aumento della potenza computazionale, il quale ha alcuni risvolti problematici. La crescita delle prestazioni dei computer ha come conseguenza di base un'inevitabile innalzamento dei costi di tali servizi e quindi una minor accessibilit alla maggioranza dei gruppi di ricerca. Questo per non l'unico effetto negativo, poich all'aumento della potenza segue un aumento del consumo energetico che non un fattore trascurabile.

Queste ragioni hanno portato, negli ultimi anni, alcuni studiosi ad interessarsi a metodi alternativi per la computazione e, in questa ricerca, approfondito il tema riguardante le simulazioni su cluster a basso consumo energetico. L'idea di fondo il poter garantire ai ricercatori un risultato, in qualit di tempi, almeno pari a quello ottenuto con la metodologia tradizionale; per con il vantaggio di consumare minor energia e spendere una quantit di fondi inferiore. Nel primo capitolo saranno introdotti e approfonditi gli elementi cardine del progetto. Il tutto a partire dall'esposizione del metodo GATK-LOD<sub>n</sub> sia nel funzionamento che nei risultati, per poi evidenziare la componente implemen-

tata in una struttura eseguibile dal tool Snakemake e, quindi, approfondire le capacità di questo strumento. Infine sarà specificato il significato di low power, così come saranno mostrate le macchine adoperate nell'analisi, per poi concludere con un accenno al funzionamento dei cluster.

Nel secondo capitolo sarà spiegato dettagliatamente il funzionamento del programma con alcuni approfondimenti sui parametri utilizzati e verranno evidenziati i passaggi per un corretto uso del sistema, dall'installazione ai dati ottenuti. In più sarà specificato quale tipo di analisi è stata compiuta su tali valori a disposizione.

Nel terzo capitolo verranno mostrati i risultati finali più rilevanti con l'aggiunta di tabelle e grafici utili ad impaginare le analisi e ad accompagnare l'esposizione. Inoltre sarà presente anche un breve accenno ai dati non considerati proficui e a coloro che sono stati trascurati.

Per terminare sarà presente un paragrafo destinato alle conclusioni, alle considerazioni finali e agli sviluppi futuri del progetto, il tutto in base agli esiti più interessanti tratti dalle indagini svolte.