



# UNIVERSITÀ DEGLI STUDI DI MILANO

**CONCORSO PUBBLICO, PER TITOLI ED ESAMI, PER IL RECLUTAMENTO DI N. 1 UNITÀ DI PERSONALE AFFERENTE ALL'AREA DELLE ELEVATE PROFESSIONALITÀ - SETTORE TECNICO-INFORMATICO, CON RAPPORTO DI LAVORO SUBORDINATO A TEMPO DETERMINATO PRESSO IL DIPARTIMENTO DI SCIENZE CLINICHE E DI COMUNITÀ NELL'AMBITO DEL PROGETTO "DIPARTIMENTI DI ECCELLENZA 2023-2027" - CODICE 22489**

La Commissione giudicatrice della selezione, nominata con Determina Direttoriale n. 14243 del 10/09/2024, composta da:

Prof.ssa Paola Causin	Presidente
Dott. Alessio Angelo Alessi	Componente
Dott. Thierry Ralph Nieux	Componente
Dott. Saverio Colangelo	Segretario

comunica le tracce relative alla prima prova.

## TEMA n. 1

Il candidato sia nella posizione di amministratore qualificato di una piattaforma HPC destinata ad utenti che svolgono ricerca in ambito bioinformatico. In relazione a tale ruolo:

1. esponga quali sono, a suo parere, le motivazioni che rendono il calcolo ad alte prestazioni, ed in particolare il calcolo parallelo, uno strumento basilare in tale campo. Individui una applicazione specifica e la porti ad esempio e supporto delle proprie affermazioni
2. discuta come organizzerebbe il supporto tecnico-scientifico per la piattaforma, discutendo strumenti e strategie
3. illustri come realizzerebbe un'architettura di calcolo capace di rappresentare una rete di unità interagenti (geni, neuroni, molecole ...). Discuta l'ottimizzazione di tale architettura e quali siano i vantaggi e limitazioni di un'implementazione in ambiente HPC.

## TEMA n. 2

Il candidato sia nella posizione di amministratore qualificato di una piattaforma HPC destinata ad utenti che svolgono ricerca in ambito bioinformatico. In relazione a tale ruolo:

1. illustri quali strumenti hardware e software possono essere usati per implementare tecniche di calcolo parallelo. Si soffermi dapprima sui modelli astratti e in seguito descriva implementazioni reali che conosce e che eventualmente ha gestito nella sua carriera lavorativa
2. spieghi come le schede GPU vengano utilizzate per accelerare i processi di calcolo e discuta quali siano i principali vantaggi delle GPU rispetto alle CPU tradizionali. Fornisca inoltre almeno un esempio di applicazioni in ambito bioinformatico che beneficiano significativamente dell'uso delle GPU
3. illustri come implementare una rete di unità interagenti (geni, neuroni, molecole...) su un calcolatore ad alte prestazioni, soffermandosi su una applicazione di sua conoscenza in questo ambito

## TEMA n. 3

Il candidato sia nella posizione di amministratore qualificato di una piattaforma HPC destinata ad utenti che svolgono ricerca in ambito bioinformatico. In relazione a tale ruolo:

1. descriva l'architettura tipica di un cluster HPC, includendo i componenti principali e il loro ruolo e descriva in particolare come possa essere gestita la comunicazione tra i nodi del cluster



2. illustri cosa si intende per data partitioning in applicazioni di calcolo parallelo. In particolare, discuta quali siano le tecniche di implementazione di tale strategia evidenziandone benefici e criticità. Individui quindi una applicazione bioinformatica e la porti ad esempio e supporto delle proprie affermazioni
3. illustri in maniera didattica come l'utilizzo del calcolo scientifico nello studio delle reti geniche ottenute da dati omici possa essere un valido ausilio alla ricerca in ambito bioinformatico

La Commissione comunica le tracce relative alla seconda prova.

## TEMA n. 1

1. Il candidato spieghi l'importanza del gestore di risorse e degli schedatori nei cluster HPC e descriva i comandi principali che un utente è solito utilizzare nell'uso quotidiano, riferendosi ad uno dei gestori che conosce (esempi: SLURM, PBS, LSF, Torque ...)
2. Un utente di un cluster HPC riferisce di aver ricevuto un errore del tipo *'stale file handle'* e lo comunica all'amministratore. Il candidato discuta quali potrebbero essere le cause di questo errore e quali passi l'amministratore del cluster dovrebbe seguire per risolverlo
3. Data una sequenza di DNA (nucleotidi A,T,C,G), il candidato produca un programma in pseudo codice che:
  - o calcoli la distanza di Hamming rispetto ad una sequenza di DNA di riferimento. Si ricorda che la distanza di Hamming corrisponde al numero di posizioni per cui le sequenze differiscono. Nel caso le due sequenze di DNA abbiano lunghezze diverse il calcolo va fatto rispetto alla sequenza più corta.

## TEMA n. 2

1. Il candidato descriva come utilizzerebbe strumenti di automazione sistemistica per installare un cluster HPC. Nel piano, includa i seguenti aspetti:
  4. Preparazione dell'ambiente: Quali passaggi iniziali svolgerebbe per preparare l'infrastruttura fisica e di rete?
  5. Installazione del sistema operativo: Come automatizzerebbe l'installazione del sistema operativo sui nodi del cluster?
  6. Configurazione dei nodi: Quali strumenti utilizzerebbe per configurare i nodi del cluster e come li impiegherebbe?
2. Un utente riferisce che in un ambiente di calcolo ad alte prestazioni che utilizza un gestore di risorse (SLURM, PBS, LSF, Torque, ...), l'esecuzione di un job genera l'errore *'Out Of Memory'*. Il candidato discuta quali potrebbero essere le cause di questo errore e quali interventi consiglierebbe all'utente per risolverlo
3. Data una sequenza di DNA (nucleotidi A,T,C,G), il candidato produca un programma in pseudo codice che:
  - o generi la sequenza inversa-complementare, che si ottiene trasformando ogni nucleotide secondo la seguente regola:
    - I. A diventa T
    - II. T diventa A
    - III. C diventa G
    - IV. G diventa C



## TEMA n. 3

1. Il candidato descriva come utilizzerebbe strumenti di automazione sistemistica per gestire un cluster HPC. Nel piano, includa i seguenti aspetti:
  - Come configurerebbe un sistema di gestione delle risorse per pianificare e monitorare i job?
  - Quali strumenti e tecniche utilizzerebbe per monitorare e mantenere il cluster in funzione?
2. Un utente riferisce che in un ambiente di calcolo ad alte prestazioni che utilizza un gestore di risorse (SLURM, PBS, LSF, Torque, ...), l'esecuzione di un job genera l'errore riportato in figura.

```
Console Terminal Background jobs >
R 4.3.2
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(igraph)
Error: package or namespace load failed for 'igraph' in dyn.load(file, DLLpath = DLLpath, ...):
unable to load shared object '/exa/software/R/4.3.2-linux-rocky8-x86_64/lib64/R/library/igraph/libs/igraph.so':
libgplk.so.40: cannot open shared object file: No such file or directory
> |
```

`library(igraph)`

Error: package or namespace load failed for 'igraph' in `dyn.load(file, DLLpath = DLLpath, ...)`:  
unable to load shared object '/exa/software/R/4.3.2-linux-rocky8-x86\_64/lib64/R/library/igraph/libs/igraph.so':  
libgplk.so.40: cannot open shared object file: No such file or directory

- Il candidato discuta quali potrebbero essere le cause di questo errore e quali interventi l'amministratore dovrebbe mettere in atto per risolverlo
3. Data una sequenza di DNA (nucleotidi A, T, C, G) e un motivo costituito da una sequenza più corta di nucleotidi, il candidato produca un programma in pseudo codice che trovi la numerosità dei motivi nella sequenza di DNA.

Milano, 1 ottobre 2024

La Commissione

Prof.ssa Paola Causin Presidente



# UNIVERSITÀ DEGLI STUDI DI MILANO

Dott. Alessio Angelo Alessi Componente

Dott. Thierry Ralph Nieuw Componente

Dott. Saverio Colangelo Segretario