CATEGORIA 1: PROGRAMMAZIONE SCIENTIFICA

DOMANDA 1:

- programmazione scientifica: librerie scientifiche

Il candidato illustri l'uso delle librerie scientifiche, i vantaggi delle stesse e nomini alcune delle piu' importanti ed utili.

CATEGORIA 1: PROGRAMMAZIONE SCIENTIFICA

ESTRATTA

DOMANDA 2:

- programmazione scientifica: debugging

Il candidato illustri i procedimenti e le tecniche piu' comuni ed eventuali strumenti associati

CATEGORIA 1: PROGRAMMAZIONE SCIENTIFICA

ESTRATTA

DOMANDA 3:

- programmazione scientifica: ottimizzazione

Il candidato illustri il concetto di ottimizzazione, alcuni metodi e tecniche ed eventuali strumenti associati

CATEGORIA 1: PROGRAMMAZIONE SCIENTIFICA

ESTRATTA

DOMANDA 4:

- programmazione scientifica: gestione dell'I/O

Il candidato illustri il ruolo dell'I/O e alcuni metodi e tecniche ed eventuali strumenti associati

CATEGORIA 1: PROGRAMMAZIONE SCIENTIFICA

ESTRATTA

DOMANDA 5:

programmazione scientifica: testing

Il candidati illustri le corrette procedure per il testing delle applicazioni scientifiche ed eventuali strumenti associati

CATEGORIA 1: PROGRAMMAZIONE SCIENTIFICA

ESTRATTA

DOMANDA 6:

programmazione scientifica: documentazione

Il candidati illustri le corrette procedure per il ~~testing~~ DOCUMENTAZIONE delle applicazioni scientifiche ed eventuali strumenti associati

CATEGORIA 2: HPC - PARALLEL COMPUTING

DOMANDA1:

Looking at the following command and its associated output, which of the following statements are FALSE

- [ ] the code runs on one single node

- [ ] the code runs in parallel

- [ ] the code is a standard benchmarks

- [ ] latency is around 1.43 microseconds

- [ ] bandwith is not calculated correctly

```
[cozzini@b13 mpi-benchmarks]$ mpirun -np 2 -npernode 1  ./IMB-MPI1 -iter 10000 -msglog 25  PingPong
#---------------------------------------------------------------
#    Intel(R) MPI Benchmarks 2019 Update 3, MPI-1 part
#---------------------------------------------------------------
# Date                  : Sat Nov  9 16:38:34 2019
# Machine               : x86_64
# System                : Linux
# Release               : 3.10.0-862.el7.x86_64
# Version               : #1 SMP Fri Apr 20 16:44:24 UTC 2018
# MPI Version           : 3.1
# MPI Thread Environment:

# Calling sequence was:
# ./IMB-MPI1 -iter 10000 -msglog 25 PingPong

#---------------------------------------------------
# Benchmarking PingPong
# #processes = 2
#---------------------------------------------------
       #bytes #repetitions      t[usec]   Mbytes/sec
            0        10000         1.23         0.00
            1        10000         1.43         0.70
            2        10000         1.49         1.34
            4        10000         1.49         2.69
```

8	10000	1.47	5.46
16	10000	1.47	10.91
32	10000	1.50	21.34
64	10000	1.57	40.64
128	10000	2.40	53.38
256	10000	2.55	100.55
512	10000	2.88	177.93
1024	10000	3.44	297.99
2048	10000	4.41	463.90
4096	10000	6.40	639.97
8192	5120	9.12	898.71
16384	2560	12.46	1314.81
32768	1280	18.03	1817.59
65536	640	29.55	2217.67
131072	320	52.84	2480.66
262144	160	94.56	2772.20
524288	80	181.02	2896.31
1048576	40	349.82	2997.44
2097152	20	692.24	3029.50
4194304	10	1377.11	3045.73
8388608	5	2834.47	2959.49
16777216	2	5698.60	2944.09
33554432	1	11433.48	2934.75

All processes entering MPI_Finalize

CATEGORIA 2: HPC - PARALLEL COMPUTING

ESTRATTA

DOMANDA 2:

Looking at the following output of the ldd command which one of the following statement are FALSE:

```bash
> ldd ./a.out
        linux-vdso.so.1 =>  (0x00007fff3cf2d000)
        libmpi.so.12 => /opt/cluster/openmpi/1.10.2/gnu/4.8.3/lib/libmpi.so.12 (0x00002ae9bdb5e000)
        libpthread.so.0 => /lib64/libpthread.so.0 (0x0000003972800000)
        libc.so.6 => /lib64/libc.so.6 (0x00000038a1a00000)
        libibverbs.so.1 => /usr/lib64/libibverbs.so.1 (0x0000003590400000)
        libopen-rte.so.12 => /opt/cluster/openmpi/1.10.2/gnu/4.8.3/lib/libopen-rte.so.12 (0x00002ae9bde4f000)
        libopen-pal.so.13 => /opt/cluster/openmpi/1.10.2/gnu/4.8.3/lib/libopen-pal.so.13 (0x00002ae9be0ca000)
        libdl.so.2 => /lib64/libdl.so.2 (0x00000038a1e00000)
        librt.so.1 => /lib64/librt.so.1 (0x0000003973400000)
        libm.so.6 => /lib64/libm.so.6 (0x00000038a2600000)
        libutil.so.1 => /lib64/libutil.so.1 (0x00000038a4e00000)
        /lib64/ld-linux-x86-64.so.2 (0x00000038a1600000)
```

- [ ] the code uses dynamic libraries
- [ ] the code is a multithreaded code
- [ ] the code is a parallel code
- [ ] the code uses static libraries

CATEGORIA 2: HPC - PARALLEL COMPUTING ESTRATTA

DOMANDA 3

A scientific code in the initialization phase reads an input file of several Gbyte. The time of this initialization phase takes 5% of the overall computation when the code runs in serial. What is the maximum number of cores you may use efficiently in a parallel run on a HPC platform?

- [ ] No more than 4 cores

- [ ] all cores you have on your multicore server that composes your HPC platform

- [ ] hard to say without knowing anything about the parallel algorithm implemented

- [ ] hard to say without knowing the cost of the core-hour on the HPC platform

ESTIMATIVA

CATEGORIA 2: HPC - PARALLEL COMPUTING

DOMANDA 4

On one node of your cluster you have user XXXXX that runs a parallel program aaaa. Using top you observe this. What is XXXXX's aaaa application likely doing?

```bash
top - 19:09:23 up 60 days,  6:44,  1 user,  load average: 8.14, 8.34, 8.50
Tasks: 787 total,   3 running, 784 sleeping,   0 stopped,   0 zombie
%Cpu(s):  0.2 us,  4.9 sy,  0.0 ni, 91.0 id,  3.6 wa,  0.0 hi,  0.2 si,  0.0 st
KiB Mem : 65754220 total, 59117108 free,  2265384 used,  4371728 buff/cache
KiB Swap:  8388604 total,  7193956 free,  1194648 used. 62963332 avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
81342 XXXXXXXX  20   0 3527312 1.043g   8132 D   3.6  1.7   1:02.66 aaaa
81338 XXXXXXXX  20   0 3527320 1.043g   8120 D   3.3  1.7   0:16.57 aaaa
81340 XXXXXXXX  20   0 3527312 1.041g   8128 D   3.3  1.7   1:01.68 aaaa
81343 XXXXXXXX  20   0 3527248 1.041g   8140 D   3.3  1.7   1:01.31 aaaa
81339 XXXXXXXX  20   0 3527248 1.043g   8004 D   2.6  1.7   1:02.20 aaaa
81341 XXXXXXXX  20   0 3527248 1.043g   8004 D   2.6  1.7   0:59.79 aaaa
81345 XXXXXXXX  20   0 3527312 1.041g   8128 D   2.6  1.7   1:02.86 aaaa
81346 XXXXXXXX  20   0 3527248 1.043g   8000 D   2.6  1.7   0:58.11 aaaa
```

- [ ] it is not doing well but we cannot understand what's wrong
- [ ] It is crunching numbers as crazy
- [ ] it is writing and or reading tons of data
- [ ] It is swapping (capital sin !)

CATEGORIA 2: HPC - PARALLEL COMPUTING

ESTRATTA

DOMANDA 5:

I submit a job to compile it and run my openMP application by means of the following script on ORFEO.

```
#PBS -l walltime=3:00:00

#PBS -l nodes=1:ppn=24


module load gnu


gcc my_openmp.c -o my_openmp.x


./my_openmp.x


exit
```

How many threads does it spawn once executed ?

- [ ] 24 threads
- [ ] 1 thread
- [ ] the compilation is wrong
- [ ] no way to know this at this stage

CATEGORIA 2: HPC - PARALLEL COMPUTING

ESTIMATA

DOMANDA 6:

You need to optimize the following piece of code.

```C
#include <stdlib.h>
#include <stdio.h>
#define SIZE 1024

int main (int arcg, char * argv[]){

 int i;
 double k = 1.232;
 double w = 5.456;
 double e_tot, y;
 fscanf( stdout, "Please insert the value of y: %.3g", y );

 for( i = 0; i < SIZE; i++){
  e_tot += ( ( k * w ) / y ) / i;
 }

 return 0;
}
```

Which of the following is surely wrong:

- [ ] compile with -O3
- [ ] compute (k*w/y) outside the loop and use just the result within the loop
- [ ] compute the inverse of i and use it within the loop
- [ ] unroll the loop

CATEGORIA 3: DISCUSSIONE PROGETTI DI PROGRAMMAZIONE ALL'INTERNO DEL LADE

ESTRATTA

Progetto 1: Gestione Immagini da SEM (Scanning Electron Microscopy)

Il laboratorio deve gestire correttamente le immagini generate dal microscopio, classificandole in base ai metadati presenti nella stessa immagine (formato .tiff) e in base ai parametri dell'esperimento. Le immagine poi devono essere processate da un motore di Machine Learning per un ulteriore classificazione sul loro contenuto. È intenzione del laboratorio creare un workflow unico per la gestione di tutte queste operazioni.

Il candidato discuta con la commissione come affronterebbe il problema.

CATEGORIA 3: DISCUSSIONE PROGETTI DI PROGRAMMAZIONE ALL'INTERNO DEL LADE

ESTRATTA

Progetto 2: Pipeline computazionali di Genomica

Il laboratorio deve, ad ogni singolo sequenziamento dello strumento Novaseq6000 effettuare il cosidetto "demultiplexing", trasformando i files dal dato proprietario Illumina al formato comune fastaq. L'operazione viene effettuata manualmente su un cluster e prevede diversi passi. È intenzione del laboratorio automatizzare completamente la procedura.

Il candidato discuta con la commissione come affrontare il problema.

CATEGORIA 3: DISCUSSIONE PROGETTI DI PROGRAMMAZIONE ALL'INTERNO DEL LADE

Progetto 3: Gestione di uno storage multi-livello.

Su un sistema HPC sono presenti piu' livelli di storage caratterizzati dalla velocita' di esecuzione e dalla dimensione. Abbiamo al momento 4 livelli: un parallel FS basato su NVME di 10TB, un parallel FS basato su SSD di 90TB, un parallel FS basato su HDD rotativi di 600TB. Infine un FS nfs di circa 1.4 PB. Gli utenti vedono tutti e quattro i FS. E' intenzione del laboratorio di gestire in maniera semi-automatica lo spostamento di files da uno all'altro dei FS, in base alla dimensione, tempi di accesso e occupazione globale dei FS.

Il candidato discuta con la commissione come affrontare il problema.

CATEGORIA 3: DISCUSSIONE PROGETTI DI PROGRAMMAZIONE ALL'INTERNO DEL LADE

ESTRATTA

Progetto 4: Misura dei consumi elettrici associati ad un datacenter

L'infrastruttura di permette di raccogliere i dati di consumo elettrico di ciascuna sua componente (server di calcolo/switches/condizionatori etc.) I dati con frequenza diversa sono tutti salvati su un DB esterno per serie temporali (Prometheus). E' intenzione del laboratorio creare una dashboard grafica per la visualizzazione dei consumi relativi alle diverse componenti dell'infrastruttura (storage/nodi di calcolo etc..)

ESTRATTA

Progetto 5: Ottimizzazione Pipelines Genomiche di Secondo Livello

I dati di genomica prodotti dal laboratorio di Genomica in formato fastaQ sono successivamente analizzati in diverse pipeline di calcolo utilizzando spesso un grande numero di software di bioinformatica. Tali software molto spesso non sono interoperabili in termini di input/output e molto spesso richiedono un approccio manuale. E' intenzione del laboratorio creare e/o utilizzare un workflow manager per gestire al meglio queste analisi.

Il candidato discuta con la commissione come affrontare il problema.

CATEGORIA 3: DISCUSSIONE PROGETTI DI PROGRAMMAZIONE ALL'INTERNO DEL LADE ESTRATTA

Progetto 6: Gestione di sofware scientifico su un'infrastruttura HPC

Il cluster e' composto da quattro tipi di nodi di calcolo: i nodi basati su architettura intel, nodi basati su architettura epyc, nodi basati su GPU V100 e nodi Nvidia DGX con GPU A100. Per sfruttare al meglio le diverse componenti hardware il software scientifico messo a disposizione del sistema deve essere compilato in maniera differente. E' intenzione del laboratorio ottimizzare le procedure di compilazione e gestione del software.

Il candidato discuta con la commissione come affrontare il problema.