# Message Passing Interface MPI

# Message Passing Interface MPI

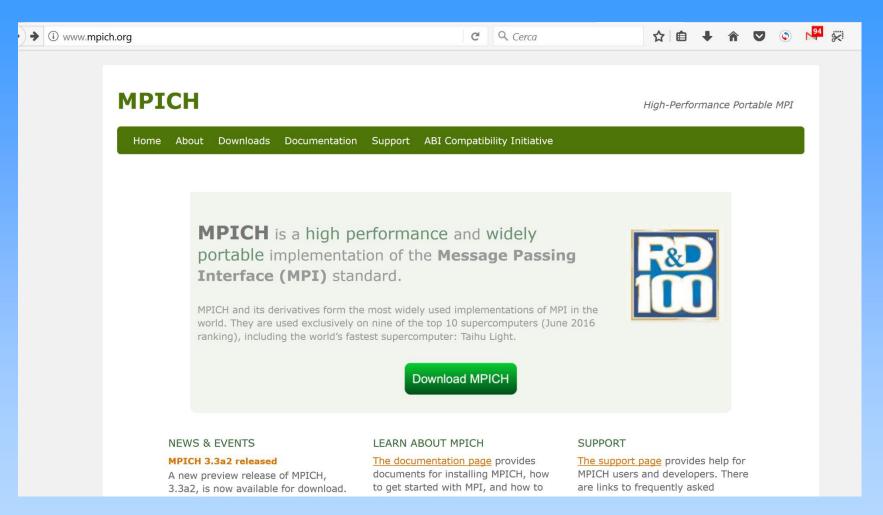
#### Implementazioni di MPI

- MPICH
  - Argonne National Laboratory (ANL) and
  - Mississippi State University (MSU)
- OpenMPI

• ..

Linguaggi ospiti:Fortran,C,C++,Matlab

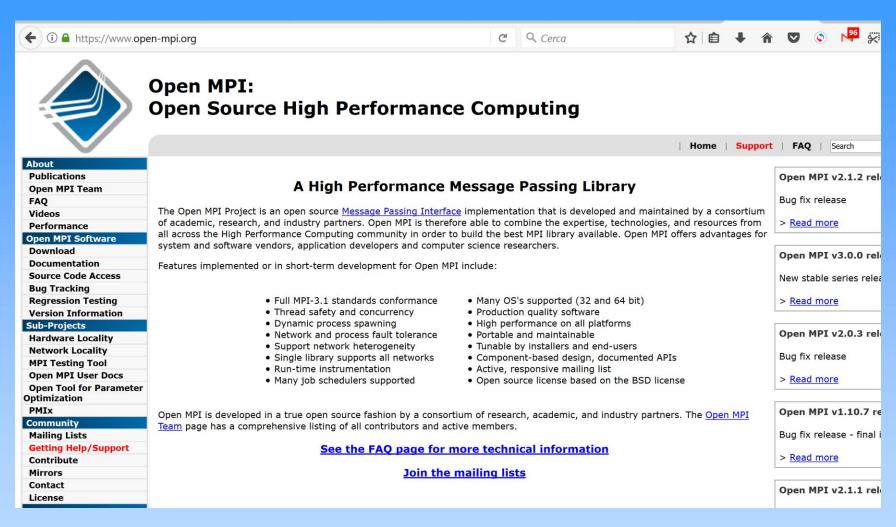
#### MPI: dove trovarlo



# http://www.mpich.org/

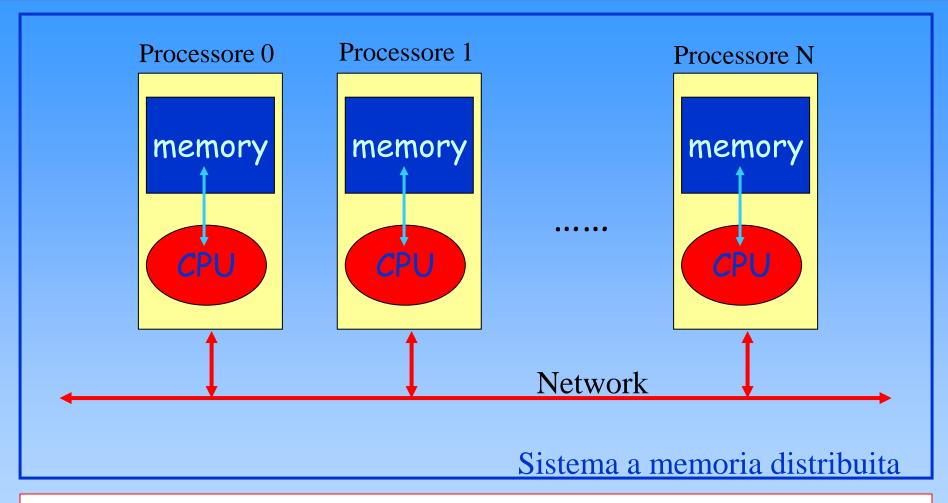
G. Laccetti - Parallel and Distributed Computing - a.a. 2023/24

#### MPI: dove trovarlo



https://www.open-mpi.org/

# Paradigma del message passing



Ogni processore ha una propria memoria locale alla quale accede direttamente.

Ogni processore può conoscere i dati in memoria di un altro processore o far conoscere i propri, attraverso il trasferimento di dati.

### Definizione: message passing

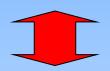
... ogni processore può conoscere i dati nella memoria di un altro processore o far conoscere i propri, attraverso il trasferimento di dati.



Message Passing:
modello per la progettazione
di software in
ambiente di Calcolo Parallelo.

#### Lo standard

La necessità di rendere portabile il modello Message Passing ha condotto alla definizione e all'implementazione di un ambiente standard.

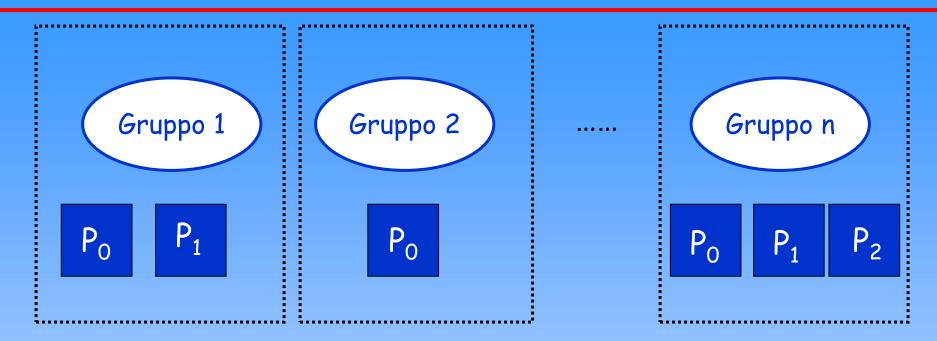


Message Passing Interface MPI

#### Per cominciare...

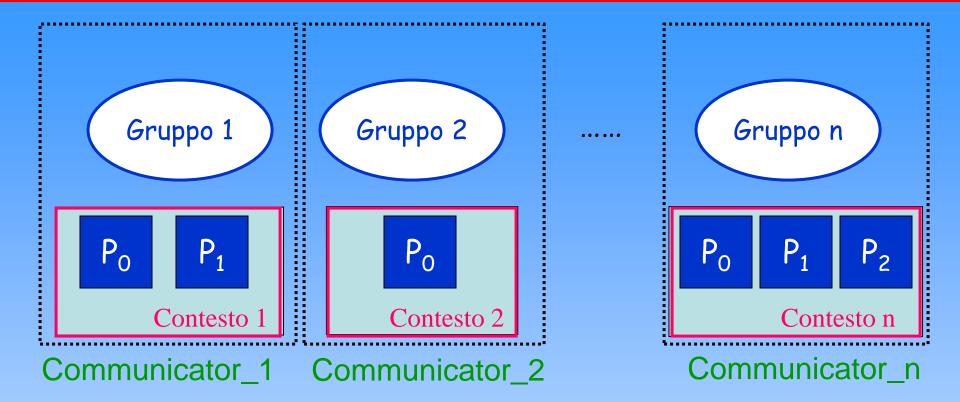
```
int main()
                                       In ambiente MPI un programma
                                           è visto come un insieme di
                                               processi concorrenti
sum=0;
for i = 0 to 14 do
sum=sum+a[i];
endfor
return 0;
main()
                                 main()
                                                                  main()
 sum=0;
                                  sum=0:
                                                                   sum=0;
 for i=0 to 4 do
                                  for i=5 to 9 do
                                                                   for i=10 to 14do
  sum=sum+a[i];
                                  sum=sum+a[i];
                                                                    sum=sum+a[i];
 endfor
                                  endfor
                                                                   endfor
return 0;
                                 return 0;
                                                                  return 0;
                  G. Laccetti - Parallel and Distributed Computing - a.a. 2023/24
```

#### Alcuni concetti di base: GRUPPO...



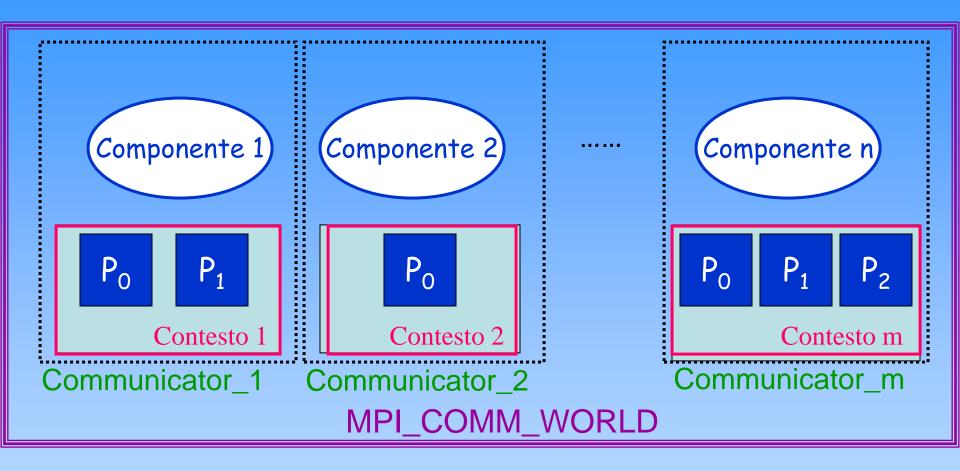
- · Ogni attività del programma viene affidata ad un gruppo di processi.
- · In ciascun gruppo ad ogni processo è associato un identificativo.
- · L'identificativo è un intero, compreso tra 0 ed il numero totale di processi appartenenti ad un gruppo decrementato di una unità.
- Processi appartenenti a uno o più gruppi possono avere identificativi diversi, ciascuno relativo ad uno specifico gruppo

#### Alcuni concetti di base: COMMUNICATOR...



- Ad un gruppo di processi appartenenti ad uno stesso contesto viene assegnato un ulteriore identificativo: il communicator.
- Il communicator racchiude tutte le caratteristiche dell'ambiente di comunicazione: topologia, quali contesti coinvolge, ecc...

#### ...Alcuni concetti di base: COMMUNICATOR



 Tutti i processi fanno parte per default di un unico communicator detto MPI\_COMM\_WORLD. MPI è una libreria che comprende:

- · Funzioni per definire l'ambiente
- · Funzioni per comunicazioni uno a uno
- Funzioni per comunicazioni collettive
- Funzioni per operazioni collettive

MPI:

# Le funzioni dell'ambiente.

# Un semplice programma:

```
#include <stdio.h>
#include "mpi.h"
int main(int argc, char *argv[])
  int menum, nproc;
  MPI_Init(&argc,&argv);
  MPI_Comm_rank(MPI_COMM_WORLD,&menum);
  MPI Comm_size(MPI_COMM_WORLD,&nproc);
  printf("Sono %d di %d\n",menum,nproc);
  MPI_Finalize();
  return 0;
```

Tutti i processi di MPI\_COMM\_WORLD stampano a video il proprio identificativo menum ed il numero di processi nproc.

G. Laccetti - Parallel and Distributed Computing - a.a. 2023/24

#include "mpi.h"

mpi.h: Header File

Il file contiene alcune direttive necessarie al preprocessore per l'utilizzo dell'MPI

```
#include <stdio.h>
#include "mpi.h"
int main(int argc, char *argv[])
  int menum, nproc;
  MPI_Init(&argc,&argv);
  MPI_Comm_rank(MPI_COMM_WORLD,&menum);
  MPI_Comm_size(MPI_COMM_WORLD,&nproc);
  printf("Sono %d di %d\n",menum,nproc);
  MPI_Finalize();
  return 0;
```

Tutti i processi di MPI\_COMM\_WORLD stampano a video il proprio identificativo menum ed il numero di processi nproc. G. Laccetti - Parallel and Distributed Computing - a.a. 2023/24



# MPI\_Init(&argc,&argv);

- · Inizializza l'ambiente di esecuzione MPI
- Inizializza il communicator MPI\_COMMON\_WORLD
- I due dati di input: argc ed argv sono gli argomenti del main

# In generale ...:

```
MPI_Init(int *argc, char ***argv);

Input: argc, **argv;
```

- Questa routine inizializza l'ambiente di esecuzione di MPI. <u>Deve essere chiamata</u> una sola volta, prima di ogni altra routine MPI.
- · Definisce l'insieme dei processi attivati (communicator).

```
#include <stdio.h>
#include "mpi.h"
int main(int argc, char *argv[])
  int menum, nproc;
  MPI_Init(&argc,&argv);
  MPI_Comm_rank(MPI_COMM_WORLD,&menum);
  MPI_Comm_size(MPI_COMM_WORLD,&nproc);
  printf("Sono %d di %d\n",menum,nproc);
  MPI_Finalize();
  return 0;
```

Tutti i processi di MPI\_COMM\_WORLD stamperanno sul video il proprio identificativo menum ed il numero di processi nproc.

G. Laccetti - Parallel and Distributed Computing - a.a. 2023/24



# MPI\_Finalize();

- Questa routine determina la fine del programma MPI.
- Dopo questa routine non è possibile richiamare nessun'altra routine di MPI.

```
#include <stdio.h>
#include "mpi.h"
int main(int argc, char *argv[])
  int menum, nproc;
  MPI_Init(&argc,&argv);
  MPI_Comm_rank(MPI_COMM_WORLD,&menum);
  MPI_Comm_size(MPI_COMM_WORLD,&nproc);
  printf("Sono %d di %d\n",menum,nproc);
  MPI_Finalize();
  return 0;
```

Tutti i processi di MPI\_COMM\_WORLD stamperanno sul video il proprio identificativo menum ed il numero di processi nproc.

G. Laccetti - Parallel and Distributed Computing - a.a. 2023/24



# MPI\_Comm\_rank(MPI\_COMM\_WORLD,&menum);

 Questa routine permette al processo chiamante, appartenente al communicator MPI\_COMM\_WORLD, di memorizzare il proprio identificativo nella variabile menum.

# In generale ...:

```
MPI_Comm_rank(MPI_Comm comm, int *menum);
```

```
<u>Input</u>: comm;
<u>Output</u>: menum.
```

 Fornisce ad ogni processo del communicator comm l'identificativo menum.

```
#include <stdio.h>
#include "mpi.h"
int main(int argc, char *argv[])
  int menum, nproc;
  MPI_Init(&argc,&argv);
  MPI_Comm_rank(MPI_COMM_WORLD,&menum);
  MPI_Comm_size(MPI_COMM_WORLD,&nproc);
  printf("Sono %d di %d\n",menum,nproc);
  MPI_Finalize();
  return 0;
```

Tutti i processi di MPI\_COMM\_WORLD stamperanno sul video il proprio identificativo menum ed il numero di processi nproc.

G. Laccetti - Parallel and Distributed Computing - a.a. 2023/24



# MPI\_Comm\_size(MPI\_COMM\_WORLD,&nproc);

 Questa routine permette al processo chiamante di memorizzare nella variabile nproc il numero totale dei processi concorrenti appartenenti al communicator MPI\_COMM\_WORLD.

# In generale ...:

```
MPI_Comm_size(MPI_Comm comm, int *nproc);
```

```
<u>Input</u>: comm;
<u>Output</u>: nproc.
```

- Ad ogni processo del communicator comm, restituisce in nproc, il numero totale di processi che costituiscono comm.
- Permette di conoscere <u>quanti processi concorrenti</u> possono essere utilizzati per una determinata operazione.

```
#!/bin/bash
  The PBS directives #
#PBS -q studenti
#PBS -I nodes=8
#PBS -N hello
#PBS -o hello.out
#PBS -e hello.err
  -q coda su cui va eseguito il job #
   -l numero di nodi richiesti #
   -N nome job(stesso del file pbs) #
   -o, -e nome files contenente l'output #
        qualche informazione sul job
NCPU=`wc -I < $PBS_NODEFILE`
echo 'This job is allocated on '${NCPU}' cpu(s)'
echo 'Job is running on node(s): '
cat $PBS NODEFILE
```

Job-script che compila ed esegue Hello.c

```
#!/bin/bash
  The PBS directives # <
#PBS -q studenti
#PBS -I nodes=8
#PBS -N Hello←
#PBS -o Hello.out
#PBS -e Hello.err
# -q coda su cui va eseguito il job #
   -l numero di nodi richiesti #
   -N nome job(stesso del file pbs) #
   -o, -e nome files contenente l'output #
       qualche informazione sul job
NCPU=`wc -I < $PBS_NODEFILE`
echo 'This job is allocated on '${NCPU}' cpu(s)'
echo 'Job is running on node(s): '
cat $PBS NODEFILE
```

La prima riga deve essere sempre questa

Le righe che iniziano con # sono commenti.
Le righe che iniziano con #PBS sono direttive.

In rosso le informazioni tipiche di questa particolare sottomissione

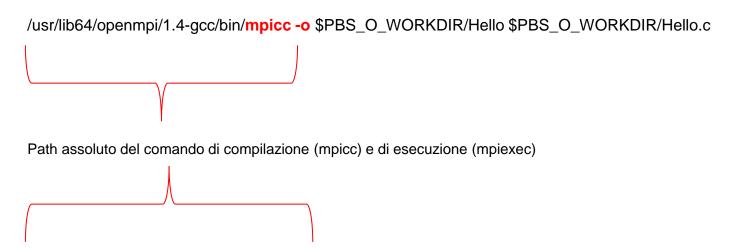
```
PBS_O_WORKDIR=$PBS_O_HOME/ProgettoHello
echo ------
echo PBS: qsub is running on $PBS_O_HOST
echo PBS: originating queue is $PBS_O_QUEUE
echo PBS: executing queue is $PBS_QUEUE
echo PBS: working directory is $PBS_O_WORKDIR
echo PBS: execution mode is $PBS_ENVIRONMENT
echo PBS: job identifier is $PBS_JOBID
echo PBS: job name is $PBS_JOBNAME
echo PBS: node file is $PBS_JOBNAME
echo PBS: current home directory is $PBS_O_HOME
echo PBS: PATH = $PBS_O_PATH
echo ------
echo "Eseguo/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o $PBS_O_WORKDIR/Hello $PBS_O_WORKDIR/Hello.c"
/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o $PBS_O_WORKDIR/Hello.c
```

echo "Eseguo:/usr/lib64/openmpi/1.4-gcc/bin/-machinefile \$PBS\_NODEFIL E -np \$NCPU \$PBS\_O\_WORKDIR/Hello" /usr/lib64/openmpi/1.4-gcc/bin/mpiexec -machinefile \$PBS\_NODEFILE \_np \$NCPU \$PBS\_O\_WORKDIR/Hello

```
PBS_O_WORKDIR=$PBS_O_HOME/ProgettoHello
echo PBS: gsub is running on $PBS O HOST
echo PBS: originating queue is $PBS O QUEUE
echo PBS: executing queue is $PBS QUEUE
echo PBS: working directory is $PBS_O_WORKDIR
echo PBS: execution mode is $PBS ENVIRONMENT
                                                             Stampa informazioni
echo PBS: job identifier is $PBS JOBID
echo PBS: job name is $PBS JOBNAME
echo PBS: node file is $PBS NODEFILE
echo PBS: current home directory is $PBS_O_HOME
echo PBS: PATH = $PBS O PATH
echo "Esequo/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o $PBS_O_WORKDIR/Hello $PBS_O_WORKDIR/Hello.c"
/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o $PB$ O WORKDIR/Hello $PB$ O WORKDIR/Hello.c
echo "Eseguo/usr/lib64/openmpi/1.4-gcc/bin/mpiexec -machinefile $PBS NODEFIL E -np $NCPU
$PBS O WORKDIR/Hello"
/usr/lib64/openmpi/1.4-gcc/bin/mpiexec -machinefile $PBS NODEFILE -np $NCPU $PBS O WORKDIR/Hello
```

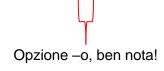
CAMBIARE IL NUMERO \$NCPU IN UNA CIFRA QUALUNQUE E FARE TUTTE LE PROVE POSSIBILI

Vediamo nel dettaglio la compilazione e l'esecuzione...



Vediamo nel dettaglio la compilazione e l'esecuzione...

/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o \$PBS\_O\_WORKDIR/Hello \$PBS\_O\_WORKDIR/Hello.c



Vediamo nel dettaglio la compilazione e l'esecuzione...

/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o \$PBS\_O\_WORKDIR/Hello \$PBS\_O\_WORKDIR/Hello.c

Nome dell'eseguibile che si vuole creare (path assoluto)

Vediamo nel dettaglio la compilazione e l'esecuzione...

/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o \$PBS\_O\_WORKDIR/Hello \$PBS\_O\_WORKDIR/Hello.c

Nome del sorgente che si vuole compilare (path assoluto)

Vediamo nel dettaglio la compilazione e l'esecuzione...

/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o \$PBS\_O\_WORKDIR/Hello \$PBS\_O\_WORKDIR/Hello.c

Elenco delle macchine utilizzabili

Vediamo nel dettaglio la compilazione e l'esecuzione...

/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o \$PBS\_O\_WORKDIR/Hello \$PBS\_O\_WORKDIR/Hello.c

Numero di processi che si devono lanciare

Vediamo nel dettaglio la compilazione e l'esecuzione...

/usr/lib64/openmpi/1.4-gcc/bin/mpicc -o \$PBS\_O\_WORKDIR/Hello \$PBS\_O\_WORKDIR/Hello.c

Eseguibile da lanciare

#### Una volta che si è scritto il file PBS

```
GLaccetti@ui-studenti:~/ProgettoHello
 NCPU= wc -1 < $PBS NODEFILE
echo 'This job is allocated on '${NCPU}' opu(s)'
echo 'Job is running on node(s): '
cat $PBS NODEFILE
PBS O WORKDIR=$PBS O HOME/HELLO
/usr/lib64/openmpi/1.2.7-gcc/bin/mpicc -o $PBS O WORKDIR/hello $PBS O WORKDIR/Hello.c
/usr/lib64/openmpi/1.2.7-gcc/bin/mpiexec -machinefile $PBS NODEFILE .np $NCPU $PBS O WORKDIR/Hello
```

Una volta che si è scritto il file PBS, e si è quindi preparata l'esecuzione parallela, si lancia con il comando

qsub Hello.pbs

qsub è in grado di interpretare le direttive contenute nel PBS

```
GLaccetti@ui-studenti ProgettoHello]$ 1s -1
total 8
-rw-r--r-- 1 GLaccetti studenti 457 Oct 19 22:55 Hello.c
-rw-r--r-- 1 GLaccetti studenti 1888 Oct 19 23:26 Hello.pbs
[GLaccetti@ui-studenti ProgettoHello]$ qsub Hello.pbs
1616.ce-studenti.scope.unina.it
[GLaccetti@ui-studenti ProgettoHello]$
```

Una volta che si è scritto il file PBS, e si è quindi preparata l'esecuzione parallela, si lancia con il comando

qsub Hello.pbs

qsub è in grado di interpretare le direttive contenute nel PBS

bisognerà aspettare che il job venga gestito dal sistema (è in coda con altri job) e che termini la sua esecuzione.

A questo punto sarà possibile visualizzare l'output e l'error con i comandi:

cat Hello.err

cat Hello.out

Per visualizzare i job: **qstat** 

Per visualizzare la coda ed il loro stato (E=eseguito,R=in esecuzione, Q= è stato accodato): qstat -q

Per eliminare un job dalla coda: **qdel jobid** 

# **Esercizi**

Scrivere, compilare ed eseguire sul cluster attraverso il PBS i seguenti esercizi. Inviare il codice, il PBS e uno o più esempi di output a valeria.mele@unina.it

- N processi. Ognuno stampa il proprio identificativo nel communicator globale
- N processi. In input: intero M>200.
   Ogni processo con identificativo >0 divide M per il proprio identificativo. Il processo 0 divide M per il numero di processi N.
  - Ogni processo stampa il risultato ottenuto.

Entro una settimana

# FINE LEZIONE