Assignment Project Exam Help

https://eduassistpro.github.i

Add WeChat edu_assist_pr

University of Leeds

Lecture 8: Introduction to distributed me

Previous lectures

Assignment Project Exam Help

In the last six lectures we looked at **shared memory parallelism** (SMP) relevant to *e.g.* multi-core CPUs:

- ory.
- https://eduassistpro.github.i
- g
- Without proper **synchronisation**, r
non-deterministic.

Add WeChat edu_assist_pr

- Dependencies can lead to **data rac**
- Can reach **deadlock** if threads wait for synchronisation events that never occur.

This lecture

Assignment Project Exam Help

This lecture is the first of six on **distributed memory parallelism**, and we will see that some (but not all) of these issues remain relev

https://eduassistpro.github.i

- ry.
- **tion**.
  - No **data races**.

Add WeChat edu_assist_pr

- Performance considerations remain th
  primary parallel overhead is **com**
- Improper synchronisation can still lead to **non-determinism** and **deadlock**.

Distributed memory systems

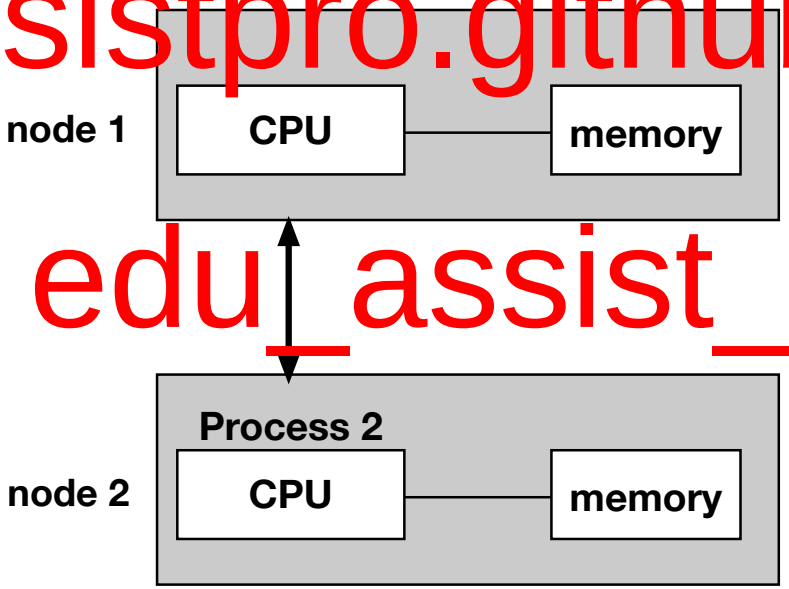
Assignment Project Exam Help

Multiple **processes** (rather than threads) that communicate *via* an **interconnection network** or **'interconnect'**.

https://eduassistpro.github.i

- F
n
- Each process has its own **heap memory**.
- If a process needs data currently held on another node's memory, must **communicate** over the network.

Add WeChat edu_assist_pr



Current fastest supercomputer[1]

Fujitsu Fugaku, RIKEN, Kobe, Japan

Assignment Project Exam Help

- ARM-based A64FX CPU.
- 4
  as
- Thttps://eduassistpro.github.i
- No GPUs.
- Draws nearly 30MW of power.Add WeChat edu_assist_pr
- Benchmarked $\approx$ 442 PFLOPS.
- 1 PFLOPS $= 10^{15}$ FLOPS.
- 1 FLOPS $=$ 1 floating point operation per second.

[1] As of Nov. 2021; `top500.org`.

Clusters as distributed systems

Assignment Project Exam Help

Supercomputers share features with other **distributed systems** such as data centres:

-
-
- May have high energy demand and cooling r

https://eduassistpro.github.i

Add WeChat edu_assist_pr

Here focus on **High Performance Co**

- Individual cluster nodes use the same **operating system**.
- Cannot usually be **addressed individually**.
- Requires a special **job scheduler**.

The interconnection network or 'interconnect'

Assignment Project Exam Help

For the local area networks within HPC clusters, communication between nodes is carried over high performance **interconnects**.

- [1].
- :
- :

https://eduassistpro.github.i

These numbers are improving with time but **CPU performance**.

Add WeChat edu_assist_pr

The need to reduce communication overhea
more important in the foreseeable future.

[1] As of Nov. 2021; see `top500.org`.

Network topology

Assignment Project Exam Help

- If data sent *via* intermediate nodes, latency is increased.
  - Each node must parse data packet and decide where to send.

Ther

Net

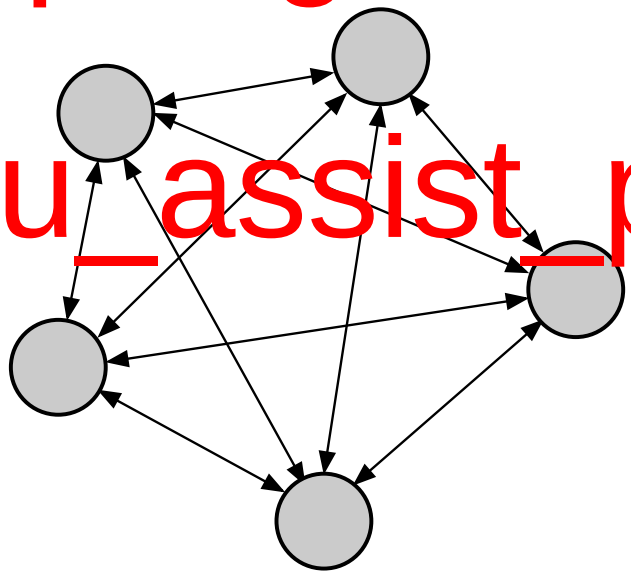
https://eduassistpro.github.i

- G = nodes *(processors)*.
- E = connections *(edges)*.

Want G with smallest **diameter** δ *(largest path length between nodes)*.

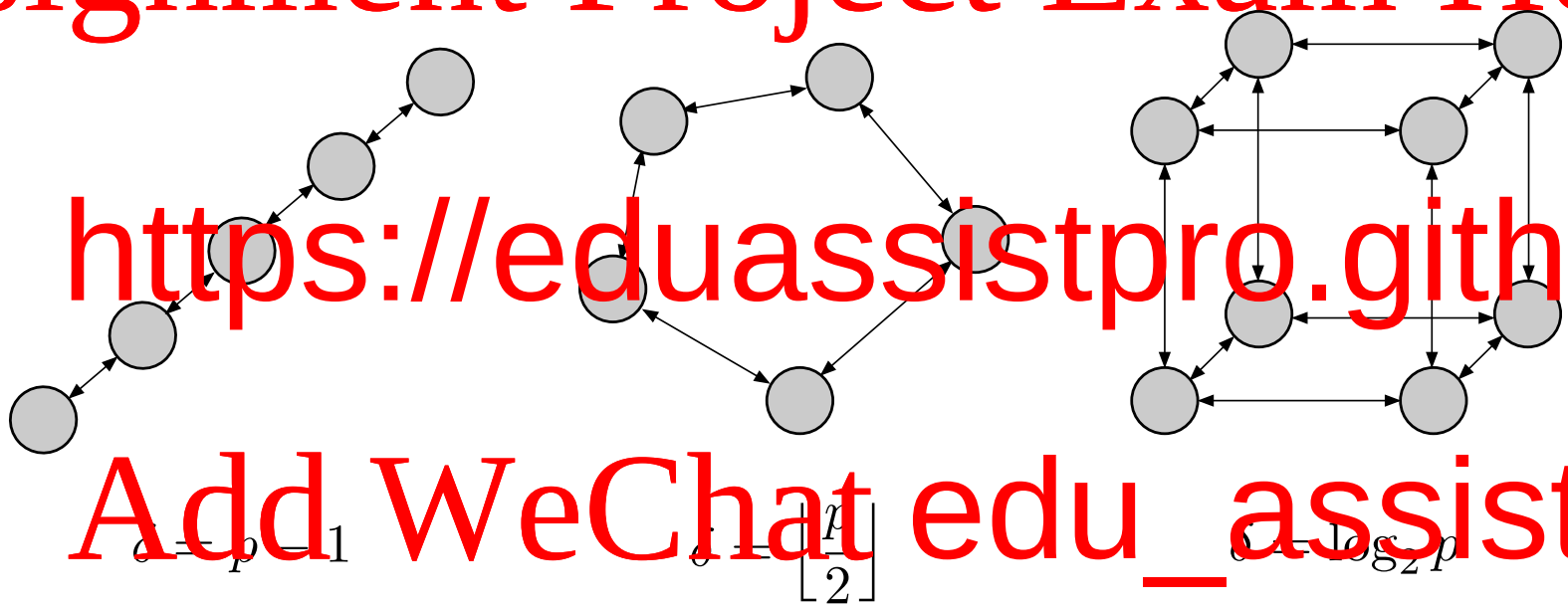
Add WeChat edu_assist_pr

A **complete graph** *(right)* has $\delta = 1$, but is impractical *(too many connections for each machine)*.



Example topologies for p nodes

Assignment Project Exam Help
https://eduassistpro.github.i
Add WeChat edu_assist_pr



Hypercube topology preferred due to its short path lengths[1].

[1] Rauber and Rünger, *Parallel programming for multicore and cluster systems* (Springer, 2013).

Processes *versus* threads

Assignment Project Exam Help

Recall from Lecture 2 that **processes** communicate with other processes using *e.g.* sockets.

https://eduassistpro.github.i

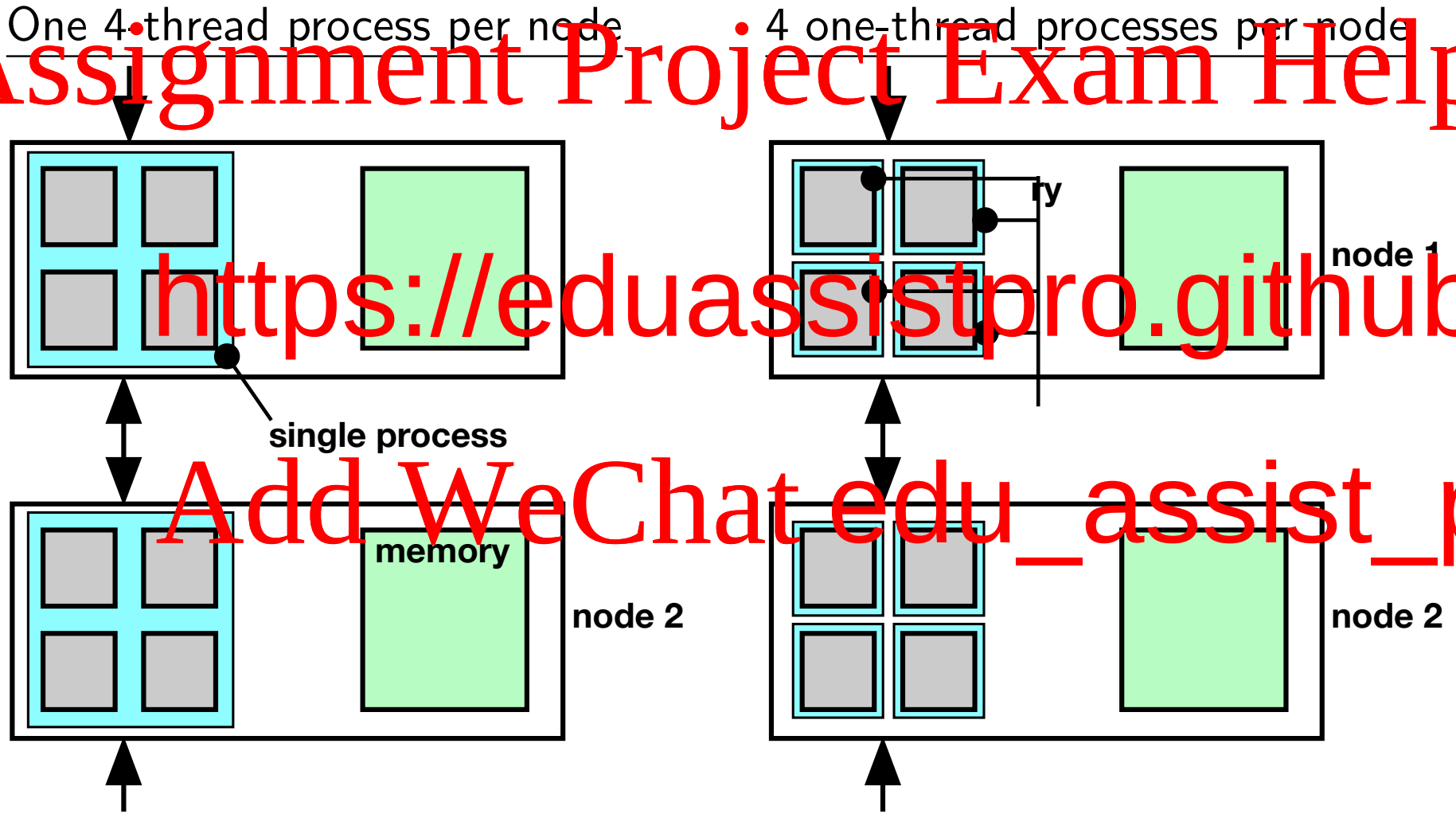
For m **ss** per node, with one thread per core.

- Avoids communication **within** a no
- Combination of OpenMP and MPI is quite c *ld*)

Add WeChat edu_assist_pr

For simplicity, we consider one **single-threaded process per core**, and therefore **multiple processes per node**.

Example for quad core nodes



Assignment Project Exam Help
https://eduassistpro.github.i
Add WeChat edu_assist_pr

Books

Assignment Project Exam Help

Wilkinson and Allen *[Lecture 1]* covers distributed memory parallelism (MPI), and a little OpenMP, but no GPU.

https://eduassistpro.github.i

A more practical book for MPI coding is:

Add WeChat edu_assist_pr

- **Parallel Programming with MPI** (Morgan-Kauffman).
  - Old (1997), only covers distributed memory systems and MPI.
  - Many code examples and snippets.

Distributed HPC programming

Assignment Project Exam Help

- For distributed HPC, there is essentially only one option[1]: **MPI**
  - Stands for **Message Passing Interface**.
  - *ing').*
  - https://eduassistpro.github.i
  - Fully supports C, C++ and FORTRAN.
    - Most online examples are in one of these lan
  - Add WeChat edu_assist_pr
  - Unofficial bindings for Java, MATLAB, P

[1] Has superseded PVM = Parallel Virtual Machine (1989). Others such as Spark, Chapel *etc.* not (yet?) widely used in HPC.

Implementations

Assignment Project Exam Help

The MP standard only defines the **interface**, it is still down to a vendor to provide an **implementation**.

https://eduassistpro.github.i

There

- **MPICH**: `www.mpich.org`
- **OpenMPI**: `www.open-mpi.org`
- Don't confuse OpenMPI with OpenMP . . . !

Add WeChat edu_assist_pr

There are also commercial implementations:

- *e.g.* Intel MPI, Spectrum MPI (IBM).

Installing MPI

Assignment Project Exam Help

The system `cloud-hpc1.leeds.ac.uk` has OpenMPI[1] installed:

```
module load mpi/openmpi-x86_64
```

https://eduassistpro.github.i

For pe
(cf. lin

- Mac users might like to try `home`

Add WeChat edu_assist_pr

On Windows machines, Microsoft MPI[2]

- Based on MPICH.

[1] Note the linux command "`module avail`" shows what modules are installed.

[2] `https://docs.microsoft.com/en-us/message-passing-interface/microsoft-mpi`

Building an MPI program

Assignment Project Exam Help

Need to use a **special compiler** for MPI programs:

- Standard installation includes `mpicc`, `mpic++`, `mpifort`.
-
- https://eduassistpro.github.i

For example, to compile a file `helloW`

Add WeChat edu_assist_pr

```
mpicc -Wall -o helloWorld helloW
```

- Will generate the executable `hel`
- All warnings on ('`-Wall`').
- Add *e.g.* `-lm` for the maths library.

Executing an MPI program

Assignment Project Exam Help

Also need a special **launcher** to execute an MPI program[1].
For multiple processes all on the **same local machine**:

```
mpiexec -n 2 ./helloWorld
```

https://eduassistpro.github.i

- `mpirun` is the same/very similar to

Add WeChat edu_assist_pr

Best to develop/debug code on a single machine (
of `cloud-hpc1.leeds.ac.uk`), then r
batch mode for *e.g.* timing runs.

[1] Executing as usual ('`./helloWorld`') will launch *one* process, *i.e.* serial.
[2] With OpenMPI, can override with the argument `-oversubscribe`.

Launching via the batch queue

Assignment Project Exam Help

The system cloud-hpc1.leeds.ac.uk has been set up to allow access to two 8-core nodes via `slurm`.

- enMP:
- https://eduassistpro.github.i

Add WeChat edu_assist_pr

```
#!/bin/bash
#Request a single node, and 8 cores (adjust as n
#SBATCH -N1 -n8

module add mpi/openmpi3-x86_64
mpiexec -n 8 ./helloWorld
```

A 'Hello World' example

```
1  #include "stdio.h"
2  #include "stdlib.h"
3  #include "mpi.h"                    // Need to include mpi.h
4
5  int main
6  {
7    int n;
8
9    MPI_Init( &argc, &argv );
10   MPI_Comm_size( MPI_COMM_WORLD, &numprocs );
11   MPI_Comm_rank( MPI_COMM_WORLD, &rank );
12
13   printf( "Process %d of %d.\n", rank, numprocs );
14
15   MPI_Finalize();
16   return EXIT_SUCCESS;
17 }
```

Assignment Project Exam Help
https://eduassistpro.github.i
Add WeChat edu_assist_pr

Initialising and finalising

Assignment Project Exam Help

The first MPI call **must** be `MPI_Init()`:

- Pass command line arguments `argc` and `argv`.
-
- https://eduassistpro.github.i

The final MPI call **must** be `MPI_Fina`

- Note the US spelling; *finalize* not

Add WeChat edu_assist_pr

Any MPI calls before `MPI_Init()` or after `MPI_Finalize()` will result in a runtime error.

Number of processes and rank

Assignment Project Exam Help

```
MPI_Comm_size(MPI_COMM_WORLD,&numprocs)
```

https://eduassistpro.github.i

```
MPI_Comm_rank(MPI_COMM_WORLD,&ra
```

Add WeChat edu_assist_pr

- Sets `rank` to the process number, know i
- Ranges from `0` to `numprocs-1` inclusive.
- Similar to `omp_get_thread_num()`.

Communicators

Assignment Project Exam Help

For our purposes, whenever you see an MPI call with the argument **communicator**, just use `MPI_COMM_WORLD`:

https://eduassistpro.github.i

In general, communicators allow processes to be

Add WeChat edu_assist_pr

- *e.g.* when developing a parallel library, to prevent processes to accidentally communicate processes.
- An advanced feature we won't consider.

Summary and next lecture

Assignment Project Exam Help

Today we have started looking at **distributed memory parallelism**:

https://eduassistpro.github.i

- For HPC, use **MPI = Message Passing Interface.**
- Seen how to build and execute a 'Hello World

Add WeChat edu_assist_pr

Next time we will see how MPI supports communication between processes, and use this to solve real problems.