分布式系统作业

第 6 次作业

姓名：唐晨轩

班级：人工智能与大数据

学号：19335182

一、 问题描述

Homework-6

一、找出给出的7个MPI程序中的bug，并将程序修改正确，比较修改前后的程序运行结果。MPI程序见附件。

二、利用MPI程序实现 n 个整形数的排序过程，排序算法不限，MPI的进程不限，可以使用单机运行多个MPI进程。提示：在排序的过程中需要MPI集合通信如MPI_Allgather等。

二、 解决方案

1.

程序 1：

rank = 0 和 rank = 1 的两个进程，进程中的 send 和 recv 函数的 tag 不一致，所以二者不能实现消息传递。

修改前：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np
4 ./mpi_bug1
Task 0 starting...
Numtasks=4. Only 2 needed. Ignoring extra...
Sent to task 1...
Task 3 starting...
Task 1 starting...
Task 2 starting...
```

修改后：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug1
Task 1 starting...
Task 3 starting...
Task 0 starting...
Numtasks=4. Only 2 needed. Ignoring extra...
Sent to task 1...
Sent to task 0...
Received from task 1...
Task 1: Received 1 char(s) from task 0 with tag 1
Task 0: Received 1 char(s) from task 1 with tag 2
Task 2 starting...
```

程序 2：

MPI_Isend 和 MPI_Irecv 的 buf_byte 不同，一个是 MPI_INT，另一个是 MPI_FLOAT。

修改前：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug2
Numtasks=4. Only 2 needed. Ignoring extra...
Task 0 sent = 0
Task 0 sent = 10
Task 0 sent = 20
Task 0 sent = 30
Task 0 sent = 40
Task 0 sent = 50
Task 0 sent = 60
Task 0 sent = 70
Task 0 sent = 80
Task 0 sent = 90
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
Task 1 received = 0.000000
```

修改后：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug2
Numtasks=4. Only 2 needed. Ignoring extra...
Task 0 sent = 0
Task 0 sent = 10
Task 0 sent = 20
Task 0 sent = 30
Task 0 sent = 40
Task 0 sent = 50
Task 0 sent = 60
Task 0 sent = 70
Task 0 sent = 80
Task 0 sent = 90
Task 1 received = 0
Task 1 received = 10
Task 1 received = 20
Task 1 received = 30
Task 1 received = 40
Task 1 received = 50
Task 1 received = 60
Task 1 received = 70
Task 1 received = 80
Task 1 received = 90
hadoop@ubuntu:~/hw6/homework6/homework6$
```

程序 3：

缺少函数 MPI_INIT 和 MPI_Finalize。

修改前：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug3
Attempting to use an MPI routine before initializing MPICH
Attempting to use an MPI routine before initializing MPICH
Attempting to use an MPI routine before initializing MPICH
Attempting to use an MPI routine before initializing MPICH
hadoop@ubuntu:~/hw6/homework6/homework6$
```

修改后：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug3
MPI task 2 has started...
MPI task 0 has started...
MPI task 1 has started...
MPI task 3 has started...
Initialized array sum = 1.335708e+14
Sent 4000000 elements to task 1 offset= 4000000
Task 1 mysum = 4.884048e+13
Sent 4000000 elements to task 2 offset= 8000000
Sent 4000000 elements to task 3 offset= 12000000
Task 2 mysum = 7.983003e+13
Task 0 mysum = 1.598859e+13
Task 3 mysum = 1.161867e+14
Sample results:
  0.000000e+00  2.000000e+00  4.000000e+00  6.000000e+00  8.000000e+00
  8.000000e+06  8.000002e+06  8.000004e+06  8.000006e+06  8.000008e+06
  1.600000e+07  1.600000e+07  1.600000e+07  1.600001e+07  1.600001e+07
  2.400000e+07  2.400000e+07  2.400000e+07  2.400001e+07  2.400001e+07
*** Final sum= 2.608458e+14 ***
hadoop@ubuntu:~/hw6/homework6/homework6$
```

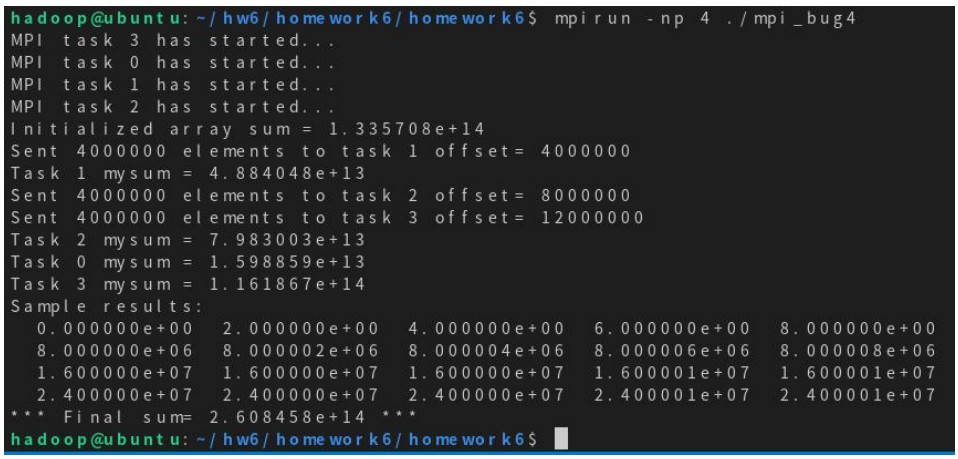
程序 4：

所有进程结束后归约操作没有用 MPI_Reduce，所以 sum 的结果不对。

修改前：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug4
MPI task 0 has started...
MPI task 3 has started...
MPI task 1 has started...
MPI task 2 has started...
Initialized array sum = 1.335708e+14
Sent 4000000 elements to task 1 offset= 4000000
Sent 4000000 elements to task 2 offset= 8000000
Task 1 mysum = 4.884048e+13
Task 2 mysum = 7.983003e+13
Sent 4000000 elements to task 3 offset= 12000000
Task 3 mysum = 1.161867e+14
Task 0 mysum = 1.598859e+13
Sample results:
  0.000000e+00  2.000000e+00  4.000000e+00  6.000000e+00  8.000000e+00
  8.000000e+06  8.000002e+06  8.000004e+06  8.000006e+06  8.000008e+06
  1.600000e+07  1.600000e+07  1.600000e+07  1.600001e+07  1.600001e+07
  2.400000e+07  2.400000e+07  2.400000e+07  2.400001e+07  2.400001e+07
*** Final sum= 1.335708e+14 ***
hadoop@ubuntu:~/hw6/homework6/homework6$
```

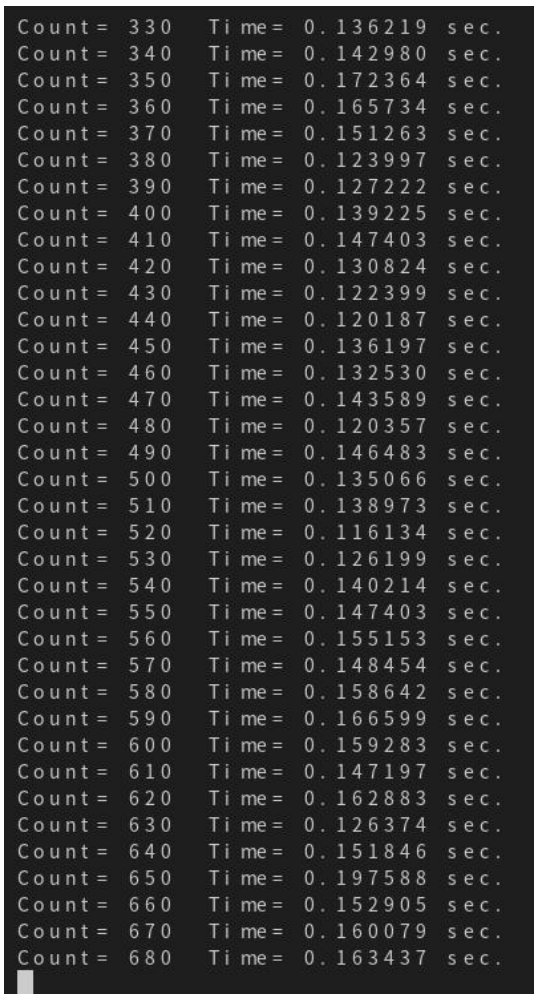
修改后：



程序 5：

由于接收消息的进程 rank1 需要花费比发送消息的进程 rank0 更多的时间，这里用的发送和接受函数均为阻塞式，所以会导致发送消息的 rank0 进入阻塞状态从而造成不合理的时间安排。

修改前：



修改后：

程序 6：

修改前只有 Isend/Irecv 的组合成功完成收发消息。因为 MPI_Waitall 函数有错误，所以 MPI_Irecv 的 buf 不能真正接收到数据。

修改前：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug6
Starting isend/irecv send/irecv test...
Task 0 starting...
Task 1 starting...
Task 2 starting...
Task 3 starting...
Task 0 has done 100 isends/irecvs
Task 0 has done 200 isends/irecvs
Task 0 has done 300 isends/irecvs
Task 0 has done 400 isends/irecvs
Task 0 has done 500 isends/irecvs
Task 0 has done 600 isends/irecvs
Task 0 has done 700 isends/irecvs
Task 0 has done 800 isends/irecvs
Task 0 has done 900 isends/irecvs
Task 0 has done 1000 isends/irecvs
Task 1 has done 100 isends/irecvs
Task 1 has done 200 isends/irecvs
Task 1 has done 300 isends/irecvs
Task 1 has done 400 isends/irecvs
Task 1 has done 500 isends/irecvs
Task 1 has done 600 isends/irecvs
Task 1 has done 700 isends/irecvs
Task 1 has done 800 isends/irecvs
Task 1 has done 900 isends/irecvs
Task 1 has done 1000 isends/irecvs
Fatal error in PMPI_Waitall: Request pending due to failure, error stack:
PMPI_Waitall(356): MPI_Waitall(count=2000, req_array=0x7fff24b14aa0, status_array=0x7fff24b169e0) failed
PMPI_Waitall(332): The supplied request in array element 0 was invalid (kind=0)
```

修改后：

程序 7：

MPI_Bcast 广播函数第二个参数有误，应该为 1

修改前：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug7
Task 1 on ubuntu starting...
Task 0 on ubuntu starting...
Root: Number of MPI tasks is: 4
Task 3 on ubuntu starting...
Task 2 on ubuntu starting...
```

修改后：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./mpi_bug7
Task 2 on ubuntu starting...
Task 3 on ubuntu starting...
Task 0 on ubuntu starting...
Root: Number of MPI tasks is: 4
Task 1 on ubuntu starting...
hadoop@ubuntu:~/hw6/homework6/homework6$
```

2.

排序算法选择了奇偶排序算法。

当仅仅使用 1 个进程的时候进入串行代码块，否则进入并行代码块：

串行时长：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpic++ oe_sort.cpp -o oe_sort
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 1 ./oe_sort
串行时间0.224789
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 1 ./oe_sort
串行时间0.223767
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 1 ./oe_sort
串行时间0.228675
hadoop@ubuntu:~/hw6/homework6/homework6$
```

4 个进程并行时长：

```
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./oe_sort
调用进程数：4 计算时间：0.217184
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./oe_sort
调用进程数：4 计算时间：0.224498
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./oe_sort
调用进程数：4 计算时间：0.174089
hadoop@ubuntu:~/hw6/homework6/homework6$ mpirun -np 4 ./oe_sort
调用进程数：4 计算时间：0.200328
hadoop@ubuntu:~/hw6/homework6/homework6$
```

排序数字个数为 9900 个，结果表现串行并行输出差异不大，我认为原因在于：

一方面输出本身比较耗时；另一方面多进程抢占输出设备资源，互相等待；

代码在压缩包中附带有。