PhastaIO Specification version 2011-11-10

PetaApps Group
Rensselaer Polytechnic Institute
Troy, New York 12180

Overview

PhataIO is a library in the Phasta CFD system for performance reading and writing of application data in "blocked" parallel approach. This document is meant to serve both as a user's guide and engineering specification document for PhastaIO. We begin with a PhastaIO example followed by a Man-page like specfication for each PhastaIO routine.

1 File Format

1.1 New file structure

The following parameters are given before writing:

1. # fields;

2. # parts in each file; (or total number of parts)

3. # files;

Note: Global part ID (GPid) associated with each part would change with each data block written and come with field tag. (But we won't have it before writing)

We will discuss later on how to pass these parameters, the options include:

1. Use the interface of current version of phastaIO.

2. Set up probably two new function to do this: one to create a string to field number mapping in writing, and the other to access this mapping information when reading.

3. The file consists of mainly two parts: `master header` and `body`. In master header, we store information like, # fields and their names, # parts in this file, # files. The most important one is the offset table which is used for looking up the field data in this file. The basic idea is illustrated by the following simple example and corresponding figure.

In the simple example below, we have 2 processors, each processor deals with 2 parts and each part contains 3 fields (e.g., solution, time derivative and error). These two processors write to one file in parallel. In multiple files case, each file is mostly the same as this example except that different processor will have different rank name. **Note: use global part ID** `GPid` **instead of rank when writing/reading parts.**.

1.2 Pseudo-code

```
2 processors: proc 0 and proc 1;
2 parts:      proc 0 deals with part 0 and part 1; Proc 1 deals with part 2 and part 3;
3 fields:     field 0, field 1 and field 2.

Call open file function: (create one file)
Setup(#fields, #files, # parts/file)
Loop over fields:  (0-2, solution, time derivative solution and error)
    Loop over parts: (proc 0: 0-1, proc 1: 2-3.)
        Compute GPID for this part data(it's changing);
        Call write header function;
        Call write datablock function;
    End loop
End loop
Close file
```

1.3 Master Header and Body Structure: 1 File Case

Below is the structure for the Master Header and Body in Figure 1 for the 2 processor, 2 parts per processor with 3 fields example from above assuming only 1 file for both processors.

The formula for finding the index into the offset table is computed as follows:

$$offset\_index \quad = \quad part_i + (field_j \times nppf) \tag{1}$$

Master Header:

MPI_IO_Tag

nFields = 3
      solution: 0
      time derivative:1
      error: 2
nParts per file = 4
nFiles = 1

Address table:
      (0, 0)
      (1, 0)
      (2, 0)
      (3, 0)
      (0, 1)
      (1, 1)
       .
       .
       .
      (3, 2)

File Body

Header:	part 0, field 0
Data:	part 0, field 0
Header:	part 1, field 0
Data:	part 1, field 0
Header:	part 2, field 0
Data:	part 2, field 0
Header:	part 3, field 0
Data:	part 3, field 0
Header:	part 0, field 1
Data:	part 0, field 1
Header:	part 1, field 1
Data:	part 1, field 1
Header:	part 2, field 1
Data:	part 2, field 1
Header:	part 3, field 1
Data:	part 3, field 1
Header:	part 0, field 2
Data:	part 0, field 2
Header:	part 1, field 2
Data:	part 1, field 2
Header:	part 2, field 2
Data:	part 2, field 2
Header:	part 3, field 2
Data:	part 3, field 2

Proc 0
Proc 1

Figure 1: Master Header and Body structure for 1 File case.

As an example, let's consider what entry $(0, 0)$ means. The first index is the global part ID, $part_i$. The second is the field

index, $field_j$. Thus, $(0, 0)$ returns the offset index into the file for $part_0$, $field_0$. $(2, 1)$ returns the offset index into the file for $part_2$, $field_1$.

Last, we plan to use @ as string separator for the field tag. That is a field tag of "solution@2" means solution field on part 2 would mean $part_i = 2$, $field_j = 0$ in the offset table of the example file, again assuming only 1 file for both processors.

Note: we decided to increase the master header size from previously defined 2MB to 4MB. This is because 2MB was too small to keep the offset data table of some cases when file number is small (i.e. file size gets bigger and offset table gets bigger too). This will not fit all cases with very small file number (e.g. 3B elements mesh with 1 file) but it will be sufficient for typical cases and we can further increase the size when we decide it is needed.

1.4 A Multi-file example

Now for a multi-file example. Consider 3 fields, 2 files, 5 parts per processor (ppp) and 4 processors (meaning a a total of 20 parts). This configuration could be read by the same configuration or by 20 processors (1 ppp, 10 processors per file), 10 processors (2 ppp, 5 processors per file), 4 processors (5 ppp, 2 processors per file). Please note, that a 5 processor configuration is not possible since while it would yield 4 parts-per-processors, we only have 2 files and so we would have an unequal distribution of parts per file which voilates one of our core assumptions. For short hand we use the following:

- number of parts per file → `nppf`

- number of parts per processor → `nppp`

Note, that even though we have shown the address table as a two dimensional array we will actually keep it as a 1 dimensional array to avoid any language-based confusion (e.g., C will layout a such a table in row major ordering where as Fortran is column major by default). The one dimensional index is easily computed from the two keys (field number and part number).

```
file 1 restart.stepnum.0              file 2 restart.<stepnum>.1
MPI_IO_TAG for File 0                 MPI_IO_TAG for File 1
nfields : 3                           nfields : 3
   solution:                             solution:
   time derivative:                      time derivative:
   error:                                error:
nParts per file (nppf): 10            nppf :    10
nFiles:   2                           nFiles :  2
(0,0)                                 (10,0)
(1,0)                                 (11,0)
(2,0)                                 (12,0)
(3,0)                                 (13,0)
(4,0)                                 (14,0)
(5,0)                                 (15,0)
(6,0)                                 (16,0)
(7,0)                                 (17,0)
(8,0)                                 (18,0)
(9,0)                                 (19,0)
(0,1)                                 (10,1)
.                                     .
.                                     .
.
(9,1)                                 (19,1)
(0,2)                                 (10,2)
.                                     .
.
.
(9,2)                                 (19,2)
HEADER(0,0)                           HEADER(10,0)
DATA(0,0)                             DATA(10,0)
```

```
.
. ORDER NOT IMPORTANT, NOR IS HAVING 0 SIZE DATA BLOCKS
.
```

We know the number of fields being written we can pre-allocate the table and access it in any $(part_i, field_j)$ order.

1.5   Application Usage, Implementation and Caveat Details

ALL THE COMMENTS BELOW ARE ON A PER FILE BASIS. THERE IS A STRONG LINK BETWEEN INIT and OPEN. ONE TO ONE AND ONTO. ACTUALLY THE ONLY REASON THEY ARE TWO SEPARATE FUNCTIONS IS TO KEEP BACKWARD COMPATIBILITY.

We add a `FinalizePHMPIO` where local comms are released and file header structure is committed. That is, a file or group of files could be closed and then re-opened and effectively append to the file group. However, once the `FinalizePHMPIO` is called, those files are now committed and we loose our ability to append to those files.

`InitPHMPIO` has a global handle as one of its arguments, to indicate which file it is working on. This handle points to the struct that is holding all the data associated with the file that we are working with. `FinalizePHMPIO` takes this handle as an argument and frees memory of the struct associated with this file. This will allow multiple files to be dealt with at the same time. (The `MAX_PHASTA_FILES` macro is used to indicate max number of files we can work on at the same time)

Other implications are as follows:

1. Application must compute total number of fields to be written before InitPHMPIO is called.

2. `InitPHMPIO`: The arguments are: `nfields`, `nppf`, `nfiles`, and returns the handle for the struct. Since we know these parameters when Phasta is started, they can be directly passed into the PhastaI/O library via this function.

3. `openfile_`: We can keep current interface where the filename can still be computed before the file open function is called. See the dicussion in the next subsection.

4. Application must concatenate the GPID to the field phrase before each field write with special separater "@".

5. Write header and data as previously described.

Another key question here is: **Upon file opening, how does a given processor know which file to open especially for the multi-file case?** The answer is the file name will contain the proper information that will enable a processor to compute which file it should open. In particular the file name format will be: $< fileTypeString > . < stepnum > . < fnum >$ where $fnum = int(mpi\_rank/nfiles)$. **That is, we will assume/require that parts are ordered and sequentially/equally distributed in ranges across ranks.**

Some notes/caveats on using as well as implementing the new PhastaIO lib that address this question are as follows:

1. On file open/create, library will create local mpicomm for parts writing to a given file. This will be set for all fields and for each "set" of parts in the multiple ppp case since we have agreed to not vary ppp across processors. Thus NO CHANGE except sequential GPID's in a given file rather than the MOD striping currently used in MPIO library.

2. Two functions to parse phrase into string, field# and GPID must be created. One for writing that "builds" the fields table and one for reading that uses the table as a map.

3. A C structure must be added to:

   - stored field string (only if new: e.g., in multiple parts-per-processor case same string will be encountered multiple times).
   - store all starting file addresses in a 64 bit integer (i.e., `long long`) in a way that the field-part locator (e.g., offset index) can be written after all parts/fields are written (when closing).
   - this field-part locator requires the order to appear sequential in the GPID index.

4. Space reserved in the header for the above field-part long-long addresses is exactly computable.

5. Currently we are planning header-then-data (sequentially one after the other). This can easily be relaxed later by storing all of the headers in a struct and, sorting them to match the table, and then writing them as a single block.

6. Note that the first GPID for a given file is $GPID_{start} = fnum * nppf$ (recall $fnum = int(mpi\_rank/nfiles)$). This can be computed for each file and correct the GPID to starting address as follows $offset_{index} = part_i + nppf * field_j$ but now $part_i = GPID_i - GPID_{start}$. Thus, the $part_i$ is not purely a global part ID anymore but is a local position relative to the file on which this processor is operating on.

The Application steps for reading a file are as follows:

1. Application must know (or use system commands to find) the number of files.

2. It will know its number of processors so, as above, each processor can determine which files hold its part's data.

3. It will not typically know the number of fields but this will be set after the file is open where it will be read

4. When application wants to read a field it will have to open the appropriate file in read mode.

5. At open time, the available fields will be loaded into the field table struct so that later strings can be mapped to field#.

6. Again the only change will be that the GPID of the part desired will have to be concatenated (by application) onto the phrase (with) for the field requested

7. A different function inside of header will then parse this phrase into the field phrase and gpid. The field phrase will be further mapped to a field number using the struct described previously.

2   queryphmpiio_

2.1   Synopsis

```
#include<phastaIO.h>

void
queryphmpiio_( const char filename[],
               int *nfields,
               int *nppf );
```

2.2   Description

This function takes filename (Michel: should we add file descriptor number as well to distinguish different files opened at same time?), open it and read the header to validate endianess and format. If endianess is correct and format is MPI new format, it will read number of fields (nfields) and number of parts per file (nppf) and pass it back to PHASTA (using a pointer).

3   initphmpiio_

3.1   Synopsis

```
#include<phastaIO.h>

int
initphmpiio_( int *nfields,
              int *nppf,
              int *nfiles,
              int *filehandle,
              char mode[]);
```

3.2 Description

This function takes the file handle and other numbers needed by header (nfields, nppf, nfiles), allocate and initialize the struct properly for this file and increases the global file descriptor index by 1. The reason we need mode in the argument is so that we need to differentiate "read" and "write". For "read" case, we presume query function was called already and MasterHeaderSize is already known; for "write" case, we need to compute a proper MasterHeaderSize we need according to our dataset size.

4 finalizephmpiio_

4.1 Synopsis

```
#include<phastaIO.h>
void
finalizephmpiio_( int *fileDescriptor );
```

4.2 Description

This function takes the file handle as argument and free the struct associated with the files pointed by file handle.

5 openfile_

5.1 Synopsis

```
#include<phastaIO.h>

void
openfile_( const char filename[],
           const char mode[],
           int*  fileDescriptor ) {
```

5.2 Description

This is the file open function. The first argument `filename[]` is a string that contains path of the file we want to open; second argument is a constant string that specifies the `mode` we want to use, i.e. "read", "write" and "append"; third argument is an integer value that contains the index of the file descriptor array(named "fileArray").
This function would push a file handle into "fileArray" and give "fileDescriptor" a proper index value.

6 readheader_

6.1 Synopsis

```
#include<phastaIO.h>

void
readheader_( int* fileDescriptor,
             const char keyphrase[],
             void* valueArray,
             int*  nItems,
             const char  datatype[],
             const char  iotype[] )
```

6.2 Description

This is the header reading function. The first argument is an `index` of file decriptor array (named "fileArray"), which specifies the right file handle; the second argument is a constant string that specifies the `field type`, e.g. "solution"; the third argument is an array that contains `types of data`(only integers so far), e.g. "nshg" and "nv"; the fourth argument is number of types of data for "valueArray"; the fifth argument is `datatype` in datablock, i.e. integer or double; the sixth argument is `iotype`, i.e. binary or non-binary.

This function reads file header and gets infomation like field type, type of data in datablock, datatype(double, integer, etc), and so on.

7 readdatablock_

```
#include<phastaIO.h>

void
readdatablock_( int*  fileDescriptor,
                const char keyphrase[],
                void* valueArray,
                int*  nItems,
                const char  datatype[],
                const char  iotype[] ) {
```

7.1 Description

This readdatablock function has the same arguments as the readheader function. The first, second, fifth, sixth arguments specify exactly the same things as readheader function. The third argument is an array that stores the data read from datablock; the fourth argument specifies the size of the data array.

This function checks field type, data type, iotype first, then uses fread(for binary data) or fscanf(for integer and double) to get data into data array.

8 writeheader_

```
#include<phastaIO.h>

void
writeheader_ (  int*  fileDescriptor,
                const char keyphrase[],
                void* valueArray,
                int* nItems,
                int* ndataItems,
                const char datatype[],
                const char iotype[]  ) {
```

8.1 Description

This writeheader function has very similar arguments as readheader function, except that it has an extra argument "ndataItems", which basically means number of data items in the datablock. This infomation is written in header right after field type.

This function gathers information like filed type, types of data(in the header such as"nshg" and "nv"), total size of data in its data block, datatype, iotype, etc, and writes these information into file header in a certain sequence.

9 writedatablock_

```
#include<phastaIO.h>
```

```
void
writedatablock_( int* fileDescriptor,
                 const char keyphrase[],
                 void* valueArray,
                 int* nItems,
                 const char datatype[],
                 const char iotype[] ) {
```

9.1   Description

This write datablock function has the same arguments as readdatablock function: the first argument is an index of file decriptor array (named "fileArray"), which specifies the right file handle; the second argument is a constant string that specifies the field type, e.g. "solution"; the third argument is an array that contains types of data(only integers so far), e.g. "nshg" and "nv"; the fourth argument is number of types of data for "valueArray"; the fifth argument is datatype in datablock, i.e. integer or double; the sixth argument is iotype, i.e. binary or non-binary.

This function checks field type, iotype, data type, and total size of data block, and then writes data into file with fwrite(for binary), or fprintf(for integer and double).

10   closefile_

```
#include<phastaIO.h>

void
closefile_( int* fileDescriptor,
            const char mode[] ) {
```

10.1   Description

This is the file close function. The first argment is the index of file descriptor array as to which file is being closed; second argument, `mode`, is an constant string which spefifies the mode, if mode is "write" or "append", a file flush function would be called before file is finally closed.

11   Limitations with using PhastaIO

What are the types of things you cannot do with PhastaIO?

12   Motivating Example

Below is an example of how PhastaIO is used.

```
#include <stdio.h>
#include <iostream>
#include <string.h>
#include <stdlib.h>
#include "mpi.h"
#include <math.h>
#include "../phastaIO.h"
#include "../rdtsc.h"

/*
 * This program uses serial read to read 1pfpp files,
 * then write out using serial write and parallel write
 * to write out files, and also do a parallel read of
```

```
 * those new files as well and finally parallel write again.
 *
 * This also covers all the user cases for testing mem
 * leaks for the library itself.
 *
 * Sample run arguments:
 * exe [starting_step_number] [num_parts] [num_output_files]
 */

#define ClockFrequency 850000000.0

using namespace std;

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

unsigned long long read_serial_timer[4];
unsigned long long write_serial_timer[4];
unsigned long long read_parallel_timer[10];
unsigned long long write_parallel_timer[10];
unsigned long long total_time[4];
float time_span[4];


read_parallel_timer[8]=rdtsc();


MPI_Init(&argc,&argv);


int myrank, numprocs;
MPI_Comm_rank(MPI_COMM_WORLD, &myrank);
MPI_Comm_size(MPI_COMM_WORLD, &numprocs);


int lstep = atoi(argv[1]); // time step number (e.g., 100 for restart.100.*)
int numparts = atoi(argv[2]); // total number of parts

int i, irstin, irstout, iarray[10], isize, nitems;
int ithree=3;
int nchar = 255;
char inrfname[255], outrfname[255], fprefix[255], fieldtag_s[255], hdata[255], path[255];
char iotype[16];
strcpy(iotype,"binary");
int magic_number = 362436;
int* mptr = &magic_number;


double **solution_field;
int *nshg_s, *nv_s, *sn_s;

// Calculate number of parts each proc deal with and where it start and end ...
int nppp = numparts/numprocs;// nppp : Number of parts per proc ...
int startpart = myrank * nppp +1;// Part id from which I (myrank) start ...
int endpart = startpart + nppp - 1;// Part id to which I (myrank) end ...

// Allocate space for each var ...
nshg_s = new int[nppp];
nv_s = new int[nppp];
sn_s = new int[nppp];
solution_field =  new double*[nppp];

// Find path of the data which you need ...
sprintf(fprefix,"./%d-procs_case",numparts);
```

```
  /***************************************************
   * start of read using serial-lib
   ***************************************************/
MPI_Barrier(MPI_COMM_WORLD);

strcpy(fieldtag_s,"solution");

// Each proc loops over parts ...
for( i = 0; i < nppp; i++ )
{

bzero((void*)inrfname,nchar);
// During each loop, find the part file ...
sprintf(inrfname,"%s/restart.%d.%d",fprefix,lstep,startpart + i);

    // this is how to time openfile ..
MPI_Barrier(MPI_COMM_WORLD);
read_serial_timer[0] = rdtsc( );
openfile_(inrfname, "read", &irstin);
MPI_Barrier(MPI_COMM_WORLD);
read_serial_timer[1] = rdtsc( );

// Read first field data ...
iarray[0]=-1;
readheader_(&irstin,fieldtag_s,(void*)iarray,&ithree,"double",iotype);
nshg_s[i]=iarray[0];
nv_s[i]=iarray[1];
sn_s[i]=iarray[2];
if(iarray[0]==-1)
{
printf( " [%d] not found ... exiting\n", myrank );
exit(1);
}
isize=nshg_s[i]*nv_s[i];
// Allocate the field data array to the right size ...
solution_field[i] = new double[isize];
// solution_field[i]=(double *) malloc( sizeof( double ) * isize );
// Now read the array ...
readdatablock_(&irstin,fieldtag_s,(void*)solution_field[i],&isize,"double",iotype);

MPI_Barrier(MPI_COMM_WORLD);
read_serial_timer[2] = rdtsc( );
closefile_(&irstin, "read");
MPI_Barrier(MPI_COMM_WORLD);
read_serial_timer[3] = rdtsc( );
}
  if(myrank ==0) cout << "********************\nSerial read finished..\n********************\n\n";

  /***************************************************
   * start of write using serial-lib
   ***************************************************/

for( i = 0; i < nppp; i++ )
{
bzero((void*)outrfname,nchar);
sprintf( path, "writeOutData", myrank+1 );
sprintf( outrfname,"%s/Serialdata.%d.%d",path, lstep, myrank+1 );

MPI_Barrier(MPI_COMM_WORLD);
```

```
write_serial_timer[0] = rdtsc( );
openfile_(outrfname, "write", &irstout);
MPI_Barrier(MPI_COMM_WORLD);
write_serial_timer[1] = rdtsc( );

    // mimic original header string
writestring_( &irstout,"# PHASTA Input File Version 2.0\n");
writestring_( &irstout, "# Byte Order Magic Number : 362436 \n");

bzero( (void*)hdata, 255 );
sprintf(hdata,"# Output generated by phPost version 2.7:  \n");
writestring_( &irstout, hdata );

isize = 1;
nitems = 1;
iarray[0] = 1;
writeheader_( &irstout, "byteorder magic number ",
(void*)iarray, &nitems, &isize, "integer", iotype );

writedatablock_( &irstout, "byteorder magic number ",
(void*)mptr, &nitems, "integer", iotype );

bzero( (void*)hdata, 255 );
sprintf(hdata,"number of modes : < 0 > %d\n", nshg_s[i]);
writestring_( &irstout, hdata );

bzero( (void*)hdata, 255 );
sprintf(hdata,"number of variables : < 0 > %d\n", nv_s[i]);
writestring_( &irstout, hdata );

isize =  nv_s[i]*nshg_s[i];
nitems = 3;
iarray[0] = nshg_s[i];
iarray[1] = nv_s[i];
iarray[2] = sn_s[i];

writeheader_( &irstout, fieldtag_s,
(void*)iarray, &nitems, &isize, "double", iotype);

nitems = isize;

writedatablock_( &irstout, fieldtag_s,
(void*)(solution_field[i]), &nitems, "double", iotype );

MPI_Barrier(MPI_COMM_WORLD);
write_serial_timer[2] = rdtsc( );
closefile_(&irstout, "write");
MPI_Barrier(MPI_COMM_WORLD);
write_serial_timer[3] = rdtsc( );
}
  if(myrank ==0) cout << "********************\nSerial write finished..\n********************\n\n";

 /****************************************************
  * start of write using parallel-lib
  ***********************************************/

/* Before writing, set up the following parameters:
   *
   * nfiles : number of files do you want to write to
```

```
 * nppf : number of parts each file has
 * nfields : number of fields we have (in this example, 2)
 * GPID : global part ID
   *
 * Thses are the key part of parallel-lib
   */


MPI_Barrier(MPI_COMM_WORLD);


int nfiles = atoi(argv[3]);
int nppf = numparts/nfiles;
int nfields = 1, GPID;



int descriptor, descriptor2;
char filename[255];
bzero((void*)filename,255);
// Procs are evenly and successively distributed to all files
// Lower-rank file will always have lower-rank procs ...
sprintf(path, "writeOutData", (int)(myrank/(numprocs/nfiles)));
sprintf(filename,"%s/regress_test_syncio_data.%d.%d",path, lstep, (int)(myrank/(numprocs/nfiles)));

  // parallel lib need to call init first
initphmpiio_(&nfields, &nppf, &nfiles,&descriptor, "write");

MPI_Barrier(MPI_COMM_WORLD);
write_parallel_timer[0] = rdtsc( );
openfile_(filename, "write", &descriptor);
MPI_Barrier(MPI_COMM_WORLD);
write_parallel_timer[1] = rdtsc( );

for (  i = 0; i < nppp; i++  )
{
GPID = myrank * nppp + i + 1;

// Write solution field ...
sprintf(fieldtag_s,"solution@%d",GPID);

isize=nshg_s[i]*nv_s[i];
iarray[0] = nshg_s[i];
iarray[1] = nv_s[i];
iarray[2] = sn_s[i];

writeheader_( &descriptor, fieldtag_s,(void*)iarray, &ithree, &isize, "double", iotype);
MPI_Barrier(MPI_COMM_WORLD);
write_parallel_timer[4] = rdtsc( );

writedatablock_( &descriptor, fieldtag_s, (void*)(solution_field[i]), &isize, "double", iotype );

}
MPI_Barrier(MPI_COMM_WORLD);
write_parallel_timer[2] = rdtsc( );
closefile_(&descriptor, "write");
MPI_Barrier(MPI_COMM_WORLD);
write_parallel_timer[3] = rdtsc( );

finalizephmpiio_(&descriptor);

  if(myrank ==0) cout << "********************\nParallel write finished..\n********************\n\n";
```

```
  /*****************************************************
   * start of read using parallel-lib
   ****************************************************/
char filename4[255];
sprintf(path, "writeOutData", (int)(myrank/(numprocs/nfiles)));
sprintf(filename4,"%s/regress_test_syncio_data.%d.%d",path, lstep, (int)(myrank/(numprocs/nfiles)));

double **new_solution_field;
new_solution_field = (double **)malloc( sizeof(double *) * nppp );
int new_iarray[3];

  queryphmpiio_(filename4, &nfields, &nppf);

initphmpiio_(&nfields, &nppf, &nfiles,&descriptor2, "read");

MPI_Barrier(MPI_COMM_WORLD);
read_parallel_timer[0] = rdtsc( );
openfile_(filename4, "read", &descriptor2);
MPI_Barrier(MPI_COMM_WORLD);
read_parallel_timer[1] = rdtsc( );

for( i = 0; i < nppp; i++ )
{
// Specify which part you want to read using GPID ...
GPID = startpart + i;
sprintf( fieldtag_s, "solution@%d", GPID );

readheader_( &descriptor2, fieldtag_s, (void*)new_iarray, &ithree, "double", iotype );
nshg_s[i]=new_iarray[0];
nv_s[i]=new_iarray[1];
sn_s[i]=new_iarray[2];
isize=nshg_s[i]*nv_s[i];
new_solution_field[i]=new double[isize];

readdatablock_( &descriptor2, fieldtag_s, (void*)(new_solution_field[i]), &isize, "double", iotype );
}

// Specify which part you want to read using GPID ...
// Always remember part id is bigger than proc id differs by 1 ...
MPI_Barrier(MPI_COMM_WORLD);
read_parallel_timer[2] = rdtsc( );
closefile_(&descriptor2, "read");
MPI_Barrier(MPI_COMM_WORLD);
read_parallel_timer[3] = rdtsc( );

finalizephmpiio_(&descriptor2);

  if(myrank ==0) cout << "********************\nParallel read finished..\n********************\n\n";

  // now let's check the value of newly read data with old posix read data
  for(i = 0; i < nppp; i++) {
    for(int j = 0; j < isize; j++) {
      if((fabs(solution_field[i][j] - new_solution_field[i][j])) > 0.000001) {
        cout << "this new value doesn't look right:\n";
        cout << "old=" << solution_field[i][j] << ", new=" << new_solution_field[i][j] << endl;
        exit(1);
      }
    }
  }
```

```
MPI_Barrier(MPI_COMM_WORLD);
   if(myrank ==0) cout << "********************\nAll field values verified..\n********************\n\n";
```

Need a detail description of what is going on this example.