

ITCS 4145/5145 Assignment 1

Using the Seeds Pattern Programming Framework

1 - Workpool Pattern

Author: B. Wilkinson and C. Ferner. Modification date: Aug. 30th, 2013

Seeds Framework

The purpose of this assignment is to become familiar with the Seeds pattern programming framework.¹ This framework was developed at UNC-Charlotte by Jeremy Villalobos as part of his PhD research into pattern programming. The framework is Java-based. The programmer first selects a particular pattern for his application. Various patterns are available including workpool, pipeline and synchronous stencil, and others can be created. We shall use the workpool pattern in this assignment as it is very general and applicable to many problems. In the workpool pattern, a master node sends out tasks to slave workers. The slaves perform computations and return results to the master, which then produces the final results. To achieve dynamic load balancing, the master keeps a queue of tasks. When slave returns a result of task, it is given another task from the queue until all tasks are completed.

To use the Seeds workpool, the programmer must implement a Java interface with three principal methods:

- The diffuse method – used by the master to distribute pieces of the data to the slaves.
- The compute method – used by the slaves for the actual computation
- The gather method – used by the master to gather the results

An additional “data count” method is used to tell the framework how many pieces of data will be computed. The programmer might implement a few other methods depending upon the application, notably an initialization method and a method to compute the final result. No message passing routines are needed by the programmer - the diffuse method will create a DataMap object with data to be passed to the slaves, and similarly the compute method will create a DataMap object with data to be passed back to the master. The framework takes care of the message passing and self deploys on a local computer, a cluster, or a geographically distributed Grid platform when the application is launched. Deployment is done using a second “bootstrapping” class with a main method. This class is mostly written for each pattern and the programmer simply fills in site-specific details (paths, etc.) prior to running.

There are three versions of the Java-based Seeds framework currently implemented:

- Full JXTA P2P networking version suitable for a fully distributed network of computers and requiring an Internet connection even in just running on a single computer.

¹ Originally the framework was called Parallel Grid Application Framework (pgaf) - “Seeds” comes from the mechanism of “seeding” computers with the framework folder during self-deployment.

- A simplified JXTA P2P version called the “NoNetwork” version for running on a single computer and not requiring an Internet connection but otherwise similar to the full JXTA P2P version.
- Multicore version implemented with threads for more efficient execution on a single multicore computer or shared memory multiprocessor. It does not require an Internet connection.

The two JXTA versions can use the same application module source code and bootstrap code, and run in the same fashion with similar logging output. The multicore version also uses the same application module source code but the bootstrap code is slightly different. In each version of the framework, only one Seeds library is different - seeds.jar, seedsNoNetwork.jar, and seedsMulticore.jar.

We will begin with the multicore version of the framework on a single PC. The Eclipse IDE will be used to provide tools to detect for coding errors. Later we will try the other versions. The assignment can be done on your own computer (recommended) or a lab computer.

Sample Code

Fully working sample programs are provided on the course home page for:

```
Monte Carlo pi ("PiApprox")
MatrixAddition ("MatrixAdd")
Matrix Multiplication ("MatrixMult")
Numerical Integration ("NumIntegration")
```

compressed into two zip files: [ProjectSource.zip](#) (network and no network versions) [ProjectSourceMulticore.zip](#) (multicore version) with the required directory structure (src.edu.uncc.grid.example.workpool....)

Project Contents:

- PiApprox:
 - MonteCarloPiModule.java
 - RunMonteCarloPi.java
- MatrixAdd:
 - MatrixAddModule.java
 - RunMatrixAddModule.java
- MatrixMult:
 - MatrixMultModule.java
 - RunMatrixMultModule.java
- NumIntegration:
 - NumericalIntegrationModule.java
 - RunNumericalIntegrationModule.java
- SeedsTemplate:
 - TemplateModule.java
 - RunTemplateModule.java

PART 1 Install Software and Execute Sample Code

Task 1: Install Java

First you need Java JDK.² To determine whether you have Java and if so, its version, type:

```
java -d64 -version
```

at the command line.³ **-d64** will establish whether you are running 32-bit Java or a 64 bit Java.

If you do not already have Java JDK installed, obtain it from:

<http://www.oracle.com/technetwork/java/javase/downloads/index.html>

32-bit and 64-bit OS. You will need to determine whether you are running a 32-bit OS or a 64 bit OS.⁴ A 32-bit OS will require a 32-bit Java.

Task 2: Install Eclipse

The Eclipse IDE will be used to build and run the Seeds code. If you do not already have Eclipse installed, obtain it from:

<http://www.eclipse.org>

Download the version of “Eclipse IDE for Java Developers” suitable for your platform and version of Java. It may take some time to download because of its size. Note the 32-bit version of Eclipse only works with 32-bit Java. The 64-bit version of Eclipse requires 64-bit OS and 64-bit Java. There is no installation wizard for Eclipse, so once you have downloaded the compressed Eclipse file, uncompress it and place the Eclipse folder in a suitable place in your file system. Eclipse can be placed anywhere for execution, but typically in C:\Program Files on a Windows XP/7 machine or /usr/local/bin on Linux. The following instructions assume a Windows system.

Eclipse is started by double clicking the Eclipse executable found in the Eclipse folder. It is convenient to create a shortcut of the Eclipse executable on the desktop.

² The assignment has been tested with JDK 1.7.0 on Windows XP and 7 computers, and JDK 1.6.0 on Mac OS X 10.6.8.

³ A Windows command line console window can be started with Win key + r, enter cmd, and press Enter. On a mac, a command line console window can be started from Applications > Utilities > Terminal.

⁴ Windows system: Start > right click (My) Computer > Properties. If you do not see x64 Edition, under system, it is a 32-bit OS. Mac: Apple Menu > About This Mac. All Mac OS X => 10.5 are 64 bit OS's.

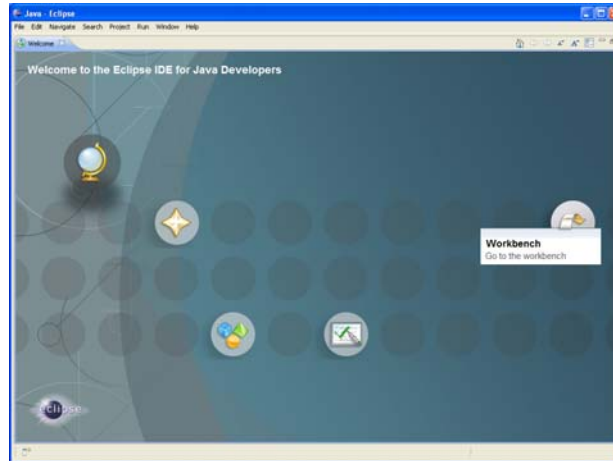


Figure 1 Welcome screen and workbench for Eclipse IDE for Java Developers

Starting the Eclipse executable will ask you the location for the workspace where your project code will be placed. At this point you will create a workspace for the assignment (Task 3 next). The appropriate Seeds software will also be placed there.

Task 3 Create Eclipse Workspace

Create a folder called **Assign1** (placed anywhere you like) and three folders within this directory:

- "workspaceMulticore" (Multicore version for operation on a single computer)
- "workspaceNoNetwork" (Version not needing an external network)
- "workspace" (for the full JXTA P2P version requiring an Internet connection)

One for each version of the framework. Once we add the Seeds software and project code later, the directory structure and important files to know are given below:

```

... --->Eclipse                               //wherever Eclipse is located
      eclipse.exe                             //click on to start Eclipse
      ...

--->Assign1
--->workspaceMultiCore                       // used to hold Seeds projects
      --->seedsMulticore                     // appropriate Seeds framework
                                              // (sometimes called pgaf)
      --->lib                                // Seeds libraries
      --->Availableservers.txt              // holds information of computers used

--->PiApprox                                 // Monte Carlo pi project
--->bin> edu>uncc>grid>example>workpool>    // Class files, empty until code compiled
--->src>edu>uncc>grid>example>workpool>    // Java source files
      MonteCarloPiModule.java              //Monte Carlo pi application code
      RunMonteCarloPiModule.java          //Bootstrap class for Monte Carlo pi code

```

```

... // other projects
--->workspaceNoNetwork // for noNetwork Seeds projects
...
--->workspace // for network version Seeds projects
...

```

Figure 2 Software directory structure

Task 4 Adding the multicore Seed libraries

We will first execute the Monte Carlo program with the multicore version, which does not require an Internet connection. Obtain the multicore version of the Seeds framework from a link “under Assignment 1 on the course home. There is no installation wizard so once you have downloaded the compressed file, uncompress it and place it in the location shown in Figure 2.

Task 5 Adding the Monte Carlo π Code

The Monte Carlo algorithm for computing π is well known and given in many parallel programming texts including the course textbook (Wilkinson and Allen 2005]. It is a so-called embarrassingly parallel application that is particularly amenable to parallel implementation but the Monte Carlo algorithm for computing π is used here more for demonstration purposes than as a good way to compute π . (It is actually a poor way to approximate π because it converges too slowly; however, the approach can lead to more important Monte Carlo applications.) A circle is formed within a square as shown in Figure 3. The circle has unit radius and the square has sides 2×2 . The ratio of the area of the circle to the area of the square is given by $\pi(1^2)/(2 \times 2) = \pi/4$. Points within the square are chosen randomly and a score is kept of how many points happen to lie within the circle. The fraction of points within the circle will be $\pi/4$, given a sufficient number of randomly selected samples. Typically only the first quadrant of the circle is used, i.e. points between 0 and 1. In the given code, the master process will send a different random number to each of the slaves. Each slave uses that number as the starting seed for their random number generator. The Java Random class `nextDouble` method returns a number uniformly

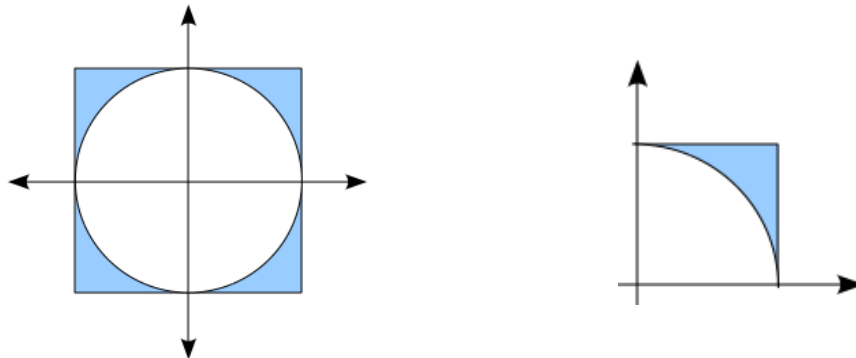


Figure 3: Unit radius circle within a 2×2 square and the upper quadrant. The ratio of the area of the circle to the area of the square is $\pi/4$. The same is true for the just the upper quadrant.

distributed between 0 and 1.0 (excluding 0 and 1). Each slave then gets the next two random numbers as the coordinates of a point (x,y) using nextDouble. If the point is within the circle (i.e. $x^2 + y^2 < 1$), it increments a counter that is counting the number of points within the circle. This is repeated for 1000 points. Each slave returns its accumulated count. The gatherData method performed by the master accumulates the slave results. A separate method, getPi, executed within the bootstrap module, computes the final approximation for π using the accumulated total.

Download Program. Download the sample source code for the Monte Carlo π workpool from the provided sample source code on the course home page. Once you have downloaded the compressed file, uncompress it and place the PiApprox folder and place it in the location shown in Figure 2. There are two Java programs:

MonteCarloPiModule.java
RunMonteCarloPiModule.java

MonteCarloPiModule.java. MonteCarloPiModule.java implements the interface for the workpool is given in Figure 4:

```
package edu.uncc.grid.example.workpool;
import java.util.Random;
import java.util.logging.Level;
import edu.uncc.grid.pgaf.datamodules.Data;
import edu.uncc.grid.pgaf.datamodules.DataMap;
import edu.uncc.grid.pgaf.interfaces.basic.Workpool;
import edu.uncc.grid.pgaf.p2p.Node;

public class MonteCarloPiModule extends Workpool {
    private static final long serialVersionUID = 1L;
    private static final int DoubleDataSize = 1000;
    double total;
    int random_samples;
    Random R;
    public MonteCarloPiModule() {
        R = new Random();
    }
    public void initializeModule(String[] args) {
        total = 0;
        Node.getLog().setLevel(Level.WARNING); // reduce verbosity for logging
        random_samples = 3000; // set number of random samples
    }
    public Data Compute (Data data) {
        DataMap<String, Object> input = (DataMap<String, Object>)data; //input gets data produced by DiffuseData()
        DataMap<String, Object> output = new DataMap<String, Object>(); // output will emit partial answers by method
        Long seed = (Long) input.get("seed"); // get random seed
        Random r = new Random();
        r.setSeed(seed);
        Long inside = 0L;
        for (int i = 0; i < DoubleDataSize ; i++) {
            double x = r.nextDouble();
            double y = r.nextDouble();
            double dist = x * x + y * y;
            if (dist <= 1.0) {
                ++inside;
            }
        }
        output.put("inside", inside); // store partial answer to return to GatherData()
        return output;
    }
    public Data DiffuseData (int segment) {
```

```

        DataMap<String, Object> d =new DataMap<String, Object>();
        d.put("seed", R.nextLong());
        return d; // returns a random seed for each job unit
    }
    public void GatherData (int segment, Data dat) {
        DataMap<String,Object> out = (DataMap<String,Object>) dat;
        Long inside = (Long) out.get("inside");
        total += inside; // aggregate answer from all the worker nodes.
    }
    public double getPi() { // returns value of pi based on the job done by all the workers
        double pi = (total / (random_samples * DoubleDataSize)) * 4;
        return pi;
    }
    public int getDataCount() {
        return random_samples;
    }
}

```

Figure 4 MonteCarloPiModule.java

In MonteCarloPiModule.java, two important classes are imported called Data and DataMap. Data is used to pass data between the master and slaves. Data is cast into DataMap within the methods and used within the methods.⁵ DiffuseData method (executed by the master) returns a random seed for each job. The Compute method (executed by slaves) picks up the random number from DiffuseData to initialize its random number generator and proceeds to choose 1000 randomly selected points counting how many are within the circle. That number is returned to the GatherData method, which accumulates all the answers from the slaves. getPi (executed in the bootstrap class) computes the final value for π .

RunMonteCarloPiModule.java. RunMonteCarloPiModule.java deploys the Seeds pattern and runs the workpool. The code for the multicore version of the framework is given in Figure 5:

```

package edu.uncc.grid.example.workpool;

import java.io.IOException;
import net.jxta.pipe.PipeID;
import edu.uncc.grid.pgaf.Anchor;
import edu.uncc.grid.pgaf.Operand;
import edu.uncc.grid.pgaf.Seeds;
import edu.uncc.grid.pgaf.p2p.Types;

public class RunMonteCarloPiModule {
    public static void main(String[] args) {
        try {
            MonteCarloPiModule pi = new MonteCarloPiModule();
            Thread id = Seeds.startPatternMulticore( new Operand( (String[])null, new Anchor( args[0],
                Types.DataFlowRole.SINK_SOURCE), pi ), 4 );
            id.join();
            System.out.println( "The result is: " + pi.getPi() );
        } catch (SecurityException e) {
            e.printStackTrace();
        } catch (IOException e) {
            e.printStackTrace();
        } catch (Exception e) {
            e.printStackTrace();
        }
    }
}

```

Figure 5 RunMonteCarloPiModule.java –Thread-based multicore version

⁵ The reason for two classes is for implementation convenience. For more details of Data and DataMap, see course note. DataMap extends Java HashMap.

Several classes are imported, PipeID, seeds-specific Anchor, Operand, Seeds, and Types. An instance of MonteCarloPiModule is first created. The Thread object is the thread managing the source and sink threads for the pattern. The programmer can monitor when the pattern is done computing by checking id.isAlive() or can just wait for the pattern to complete using id.join(). Args[0] should be the local host name.

The Seeds method startPatternMulticore starts the workpool pattern on the computer. It requires as a single argument an Operand object. Creating an Operand object requires three arguments. The first is a String list of argument that will be submitted to the host. The second is an Anchor object specifying the nodes that should have source and sink nodes (the master in this case) which in this provided as the string argument of main (first command line argument, args[0]). The third argument is an instance of MonteCarloPiModule previously created. As mentioned above, to run this code, we will need to provide one command line argument, the name of the local host.

Task 6 Executing Monte Carlo π Code on the multicore version of Seeds

The program can be executed on the command line or through an IDE. We choose here to use Eclipse.

Step 1. Open Eclipse

Start Eclipse and go to the workbench. Select the workspace you created called **workspaceMulticore**.

Step 2 Create a new project

Go to **File > New > Project** and select **Java Project**, or **File > New > Java Project**. Provide a name for the project, say PiApprox. If you click **Finish** at this time, you will see a new project created on the left panel as shown on the right image of Figure 6.

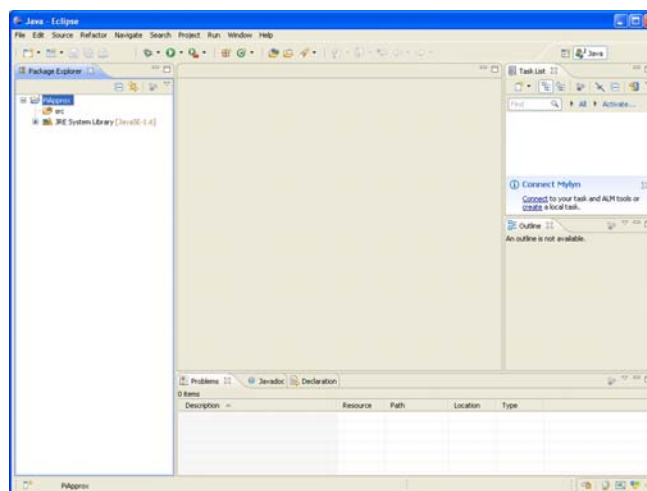


Figure 6 Creating the PiApprox Java project

Step 3 Add source code

At this point you should already see the two source files **MonteCarloPiModule.java** and **RunMonteCarloPiModule.java** within **src/edu/uncc/grid/example/workpool**, as shown in Figure 7. Select **File > Refresh** if needed. If you do not see the src files, you will need to add them, by for example dragging and dropping the source file directory into the project navigator window replacing the existing src directory. In any event, you will have unresolved errors as we have not yet added the paths to the libraries.

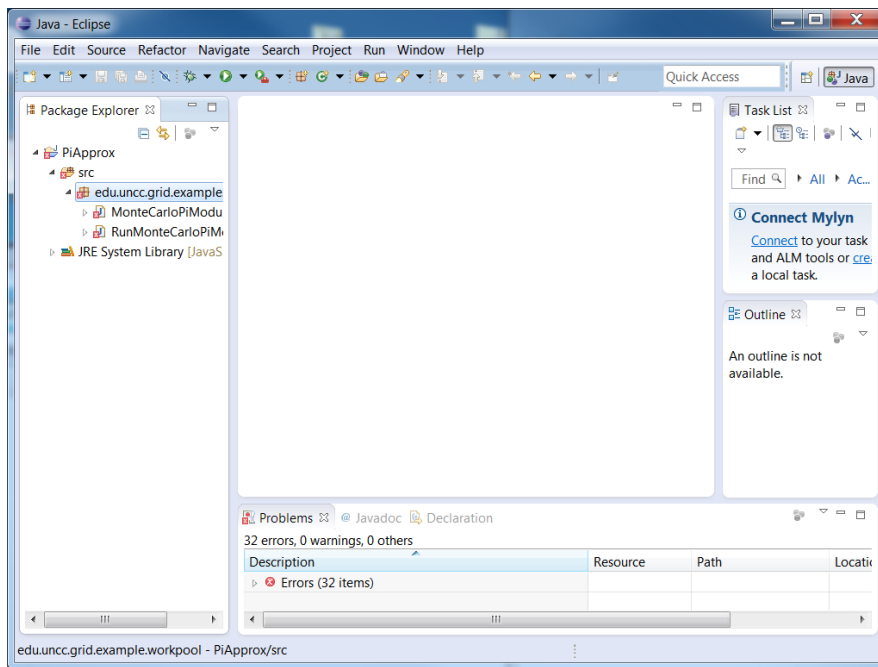


Figure 7 Source files

Step 4 Add build path to Seeds libraries

Right click the **PiApprox** folder and select **Properties**. Select **Java Build Path > Libraries tab > Add library > User Library** and **Next**. Since you have not yet provided any named user libraries you will not see any user libraries. Select **User Library** again and **New** and provide a name for the libraries, in this case say “SeedsMulticore” and click **OK**. Click on **Add External Jars** and navigate through your file system to the **Assign1/workspaceMulticore/seedsMulticore/lib** folder. Select all the jars inside lib (**control-A**). Click **Open**. Finally you should see something like Figure 8. Click **OK**, and get back to the workbench (**Finish > OK**). At this point, all unresolved references should vanish.

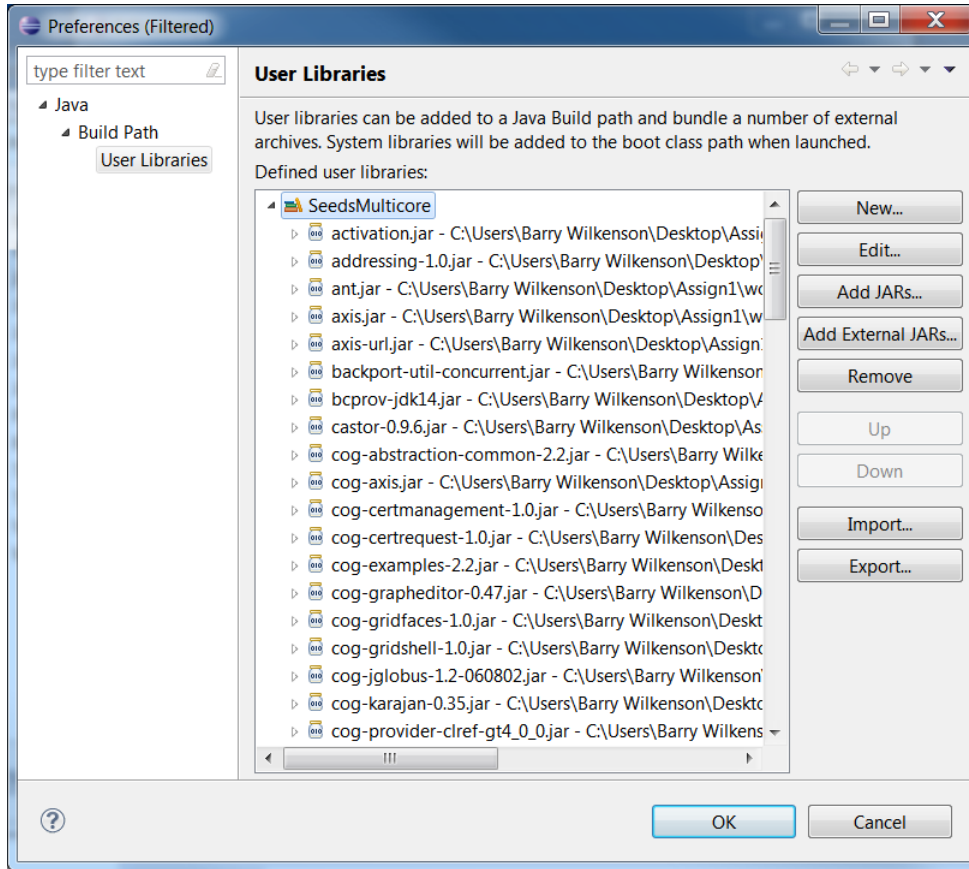


Figure 8 Seeds libraries in build path

For subsequent Seeds projects, you will be able to simply select the named Seeds libraries. Note each of the three versions of Seeds have different Seeds libraries and should be named differently.

Step 5 Command Line Arguments

Before you can run the program, you will need to provide a command line argument that is read by the bootstrap class **RunMonteCarloPiModule**. Go to **Run > Run Configurations > Java Application** (Figure 9).

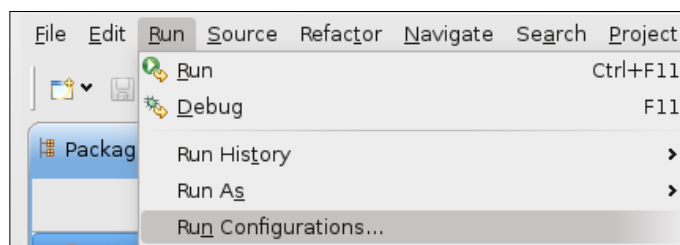


Figure 9 Run configurations

A Java Application configuration called **RunMonteCarloPiModule** should already be present. If so, rename it to **PiApprox**, otherwise create a named configuration by clicking on the leftmost icon for **New launch configuration** at the top left of **Run Configurations** to create a new configuration named **PiApprox**. Select the **PiApprox** configuration.

Main class. In the main tab, confirm the main class is **edu.uncc.grid.example.workpool.RunMonteCarloPiModule**, otherwise set the main class to that.

Arguments. Click the tab named **(x)=Arguments**. For the multicore version of Seeds, the bootstrap class is written to accept one argument, the name of your computer. The name of your computer can be found by typing hostname on the command line. Enter this name as the program argument (shown as <computerName> in Figure 10).

IMPORTANT. Do NOT use the computer name that you will see from "View system information" or similar, which can have additional characters added to the name. You MUST use the name returned by the hostname command.

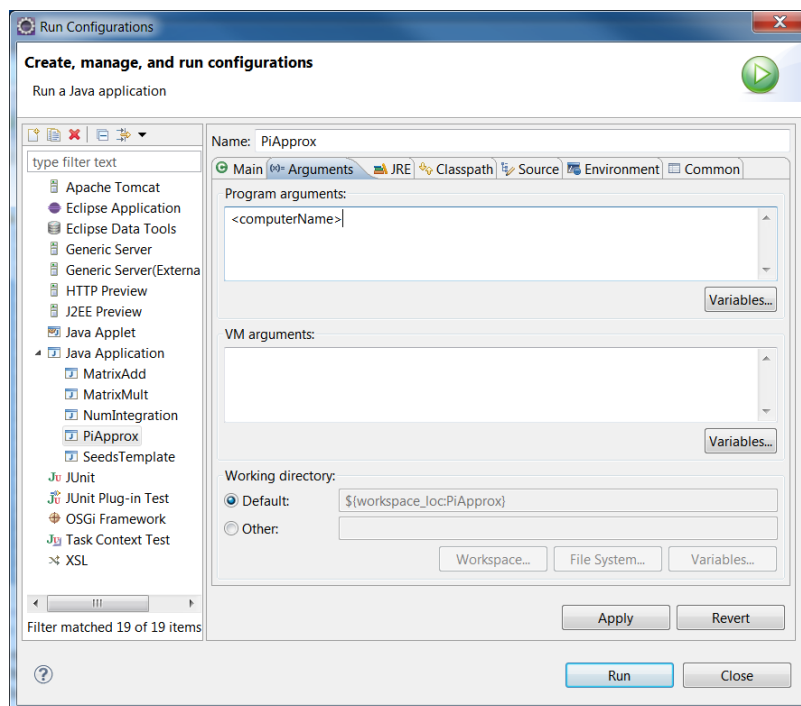


Figure 10 Program command line argument

Step 6 Run program

Click "Run" to run the project. You should see the project run immediately with output in the console window (Figure 11).

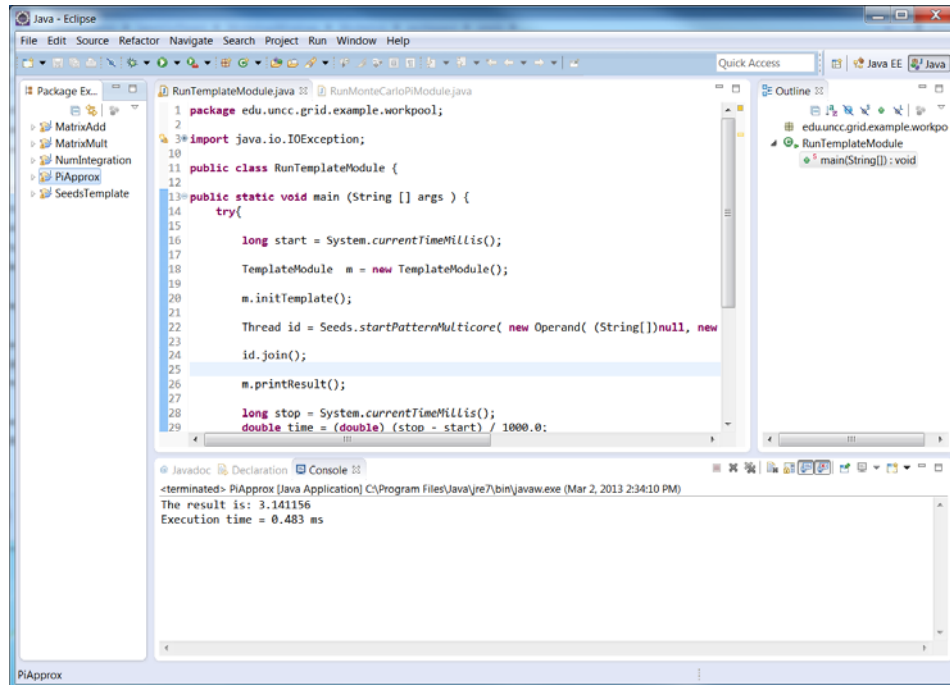


Figure 11 PiApprox output

How many random numbers were tried by the π approximation program?

Task 7 Correcting a Flaw in Monte Carlo π Code

There is a flaw in the Monte Carlo π code. Although it produces the correct answer, the use of a random number to start each random number sequence in each slave using the same random function causes each sequence to be interrelated. Modify the code to fix this problem and execute the code. Provide the code, a full explanation, and results in your write up.

Task 8 Matrix Addition and Multiplication

Repeat the process for running PiApprox to execute the provided sample programs for matrix addition and matrix multiplication and report on the results. Test the code with the following matrices:

Matrix A

```
1 2 3
4 5 6
7 8 9
```

Matrix B

```
9 8 7
6 5 4
3 2 1
```

Make sure you show the result that you get in your write up as a screenshot.

What to submit for Part 1

Your submission document should include the following:

- 1) Your name and school;
- 2) Whether you are a graduate or undergraduate student;
- 3) Screenshot of the running Monte Carlo π program and the output.
- 4) How you corrected the flaw in the Monte Carlo π code, a listing your code (*not as screen shots*) with an explanation, and the execution results as a screen shot.
- 5) Screenshot of the running matrix addition program and the output.
- 6) Screenshot of the running matrix multiplication program and the output.

PART 2 NETWORK VERSION OF SEEDS FRAMEWORK

Now we will try the JXTA P2P versions of Seeds, which are intended for a distributed computer platform. There are two JXTA P2P versions of Seeds. The full network version of Seeds requires an Internet connection. The “NoNetwork” JXTA P2P version is a similar JXTA P2P implementation but runs on a single computer without an Internet connection. For this section, you can use either of the JXPA P2P versions. Obviously if you do not have an Internet connection, use the “NoNetwork” version. In that case, choose "workspaceNoNetwork" workspace rather than “workspace” workspace in the following.

RunMonteCarloPiModule.java. RunMonteCarloPiModule.java deploys the Seeds pattern and runs the workpool. The code for the network version of the framework is given in Figure 12:

```
package edu.uncc.grid.example.workpool;

import java.io.IOException;

import net.jxta.pipe.PipeID;
import edu.uncc.grid.pgaf.Anchor;
import edu.uncc.grid.pgaf.Operand;
import edu.uncc.grid.pgaf.Seeds;
import edu.uncc.grid.pgaf.p2p.Types;

public class RunMonteCarloPiModule {

    public static void main(String[] args) {
        try {
            MonteCarloPiModule pi = new MonteCarloPiModule();
            Seeds.start( args[0] , false);

            PipeID id = Seeds.startPattern(
                new Operand( (String[])null, new Anchor( args[1] , Types.DataFlowRoll.SINK_SOURCE), pi ));
            System.out.println(id.toString() );
            Seeds.waitOnPattern(id);
            System.out.println( "The result is: " + pi.getPi() );

            Seeds.stop();

        } catch (SecurityException e) {
            e.printStackTrace();
        }
    }
}
```

```

    } catch (IOException e) {
        e.printStackTrace();
    } catch (Exception e) {
        e.printStackTrace();
    }
}
}
}

```

Figure 12 RunMonteCarloPiModule.java – Network version

The code is quite similar to the multicore version except now Seeds is started and deployed on the list servers using the Seeds method start, which takes as its first argument the path to the seeds folder on the local computer. In the code given, the path is provided as the string argument of main (first command line argument, args[0]). The Seeds method startPattern starts the workpool pattern on the computers. It requires as a single argument an Operand object. Creating an Operand object requires three arguments. The first is a String list of argument that will be submitted to the remote hosts. The second is an Anchor object specifying the nodes that should have source and sink nodes (the master in this case) which in this provided as the string argument of main (second command line argument, args[1]). The third argument is an instance of MonteCarloPiModule previously created. The Seeds method waitOn Pattern waits for the pattern to complete, after which the results are obtained using the getPi method in MonteCarloPiModule.java. Seeds is stopped using the method stop.

As mentioned, to run this code, we will need to provide two command line arguments, the local path to the Seeds folder and the name of the local host. Both could have been hardcoded.

It is now necessary to specify the servers, even though in this case we will only use a single computer.

Task 1 Specifying the computers to use

The **AvailableServers.txt** file found inside the **seeds** folder within the workspace folder needs to hold the name of the computers being used and other information can be included. For this session, we will only use a local computer and just need to provide its name of the computer. Lines starting with a # are commented out lines. Modify the one uncommented line:

```
<computerName> local - - - 1 10 GridTwo
```

replacing <computerName> (or whatever name is there) with the name of your computer and set the number of processors from 1 to however many processors you have (normally just one) and set the number of cores from 10 to the number of cores in each processor on your computer.

The name of your computer can be found by typing hostname on the command line.

Do NOT use the computer name that you will see from "View system information" or similar, which can have additional characters added to the name. You MUST use the name returned by the hostname command.

Task 2 Executing the Monte Carlo π program.

Steps 1-4 Open Eclipse. Create a new project. Add source code. Add build path to Seeds libraries

Follow the same steps as previously to create a PiApprox project but for the JXTA P2P version of Seeds. Note the Internet version of Seeds has different libraries to the multicore version and should be given unique name, say Seeds for the network version or SeedsNoNetwork for the no network version.

Step 5 Command Line Arguments

For the full network version and “NoNetwork” version of Seeds, the bootstrap class is written to accept two arguments:

1st argument: Path to where AvailableServers.txt is located

2nd argument: Name of your computer

Enter the two arguments (Figure 13). Include double quotes to make a string if there are one or more spaces in the path. The name of your computer should be the same as you put in AvailableServers.txt.

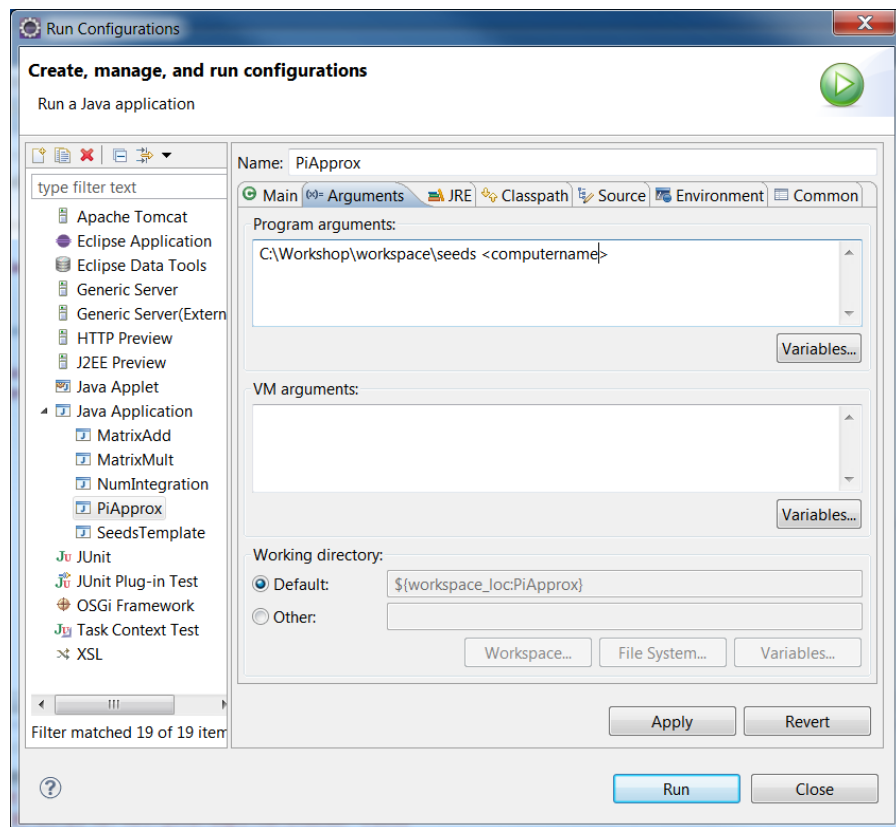


Figure 13 Program command line arguments

Step 4 Run program

Click “Run” to run the project. The console output will begin with logging messages such as Figure 14.

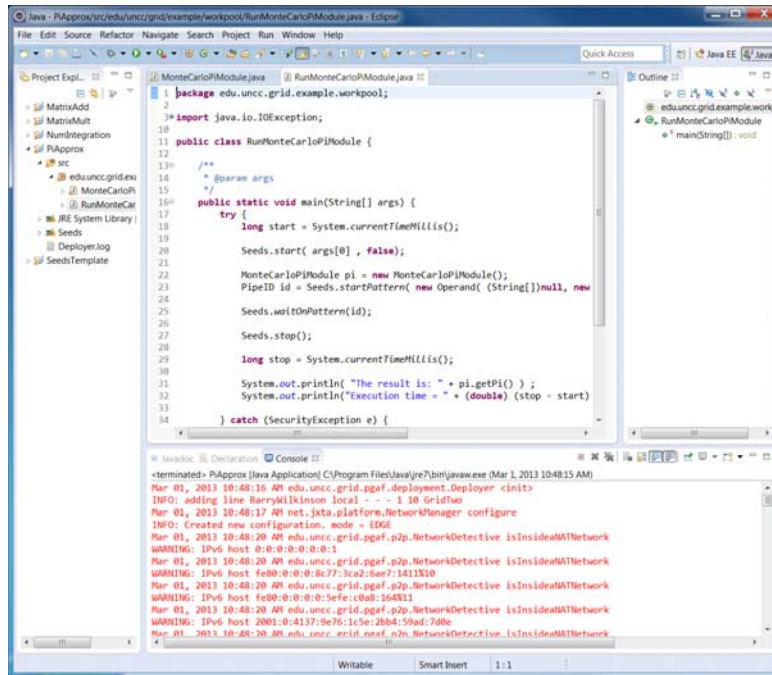


Figure 14 Sample PiApprox logging messages (Network versions)

The final result is in black at the end of the messages as shown in Figure 15:

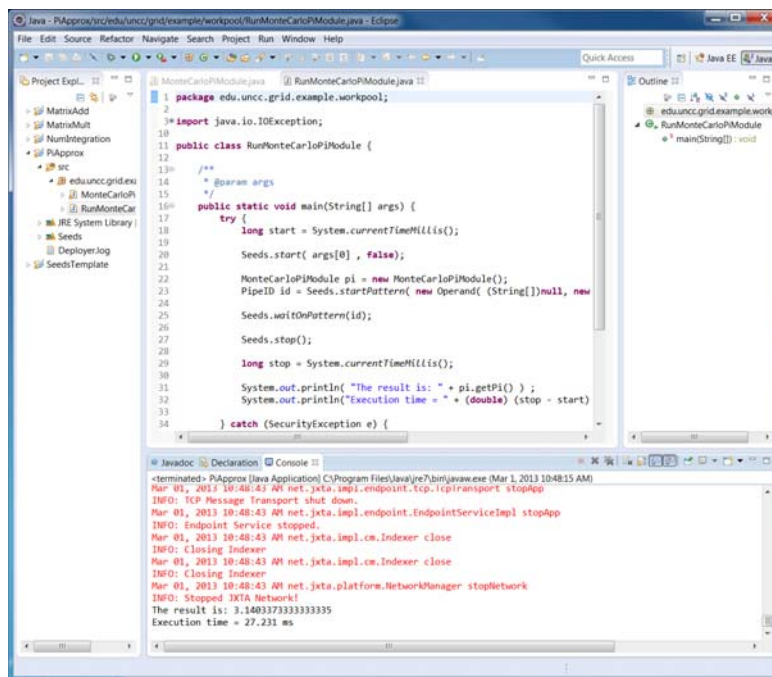


Figure 15 Sample PiApprox result (Network versions)

Issues running program: If you do not get the expected output, see posted FAQs on the course home page for known issues.

What to submit for Part 2

Your submission document should include the following:

- 1) Screenshot of the running your Monte Carlo π program, some of the red output, and the final black output given the value for π , using one of the network versions of Seeds.

PART 3 Writing Your Own Code

Write a workpool Seeds program to transpose a 5 x 5 matrix. Include measurement of time of execution. Demonstrate your program with the matrix:

Matrix

```
1 2 3 6 9
4 5 6 3 8
7 8 9 2 1
2 5 2 3 8
9 4 9 2 5
```

Execute the program on the multicore version of Seeds.

What to submit for Part 3

Your submission document should include the following:

- 1) Your matrix transpose program with comments (*not as screen shots*)
- 2) Screenshot of the running your matrix transpose program on the multicore version of Seeds. Clearly show the output.

Your submission document should include insightful conclusions.

Grading

Every part and task specified will be allocated a score so make sure you clearly identify each part and task you did.

Assignment Submission

Submit **ONE PDF** file containing everything on the UNC-C Moodle-2 by the due date as described on the course home page.