HowToDebugBSPPrograms

How to debug your own BSP applications

Debugging distributed programs is always difficult, because very few debuggers will let you connect to a remote program that wasn't run with the proper command line arguments.

1. Start by getting everything running (likely on a small input) in the local runner.
   You do this by setting your BSP Master to "local" in your config. The local runner can run
   under the debugger and runs on your development machine. A very quick and easy way to set this
   config variable is to include the following line just before you run the job:

```
conf.set("bsp.master.address", "local");
```

Running in local mode makes the job run within 20 threads by default. Since this isn't always very convenient to debug, you can decrease the number of tasks with this line:

```
conf.set("bsp.local.tasks.maximum", "2")
```

Obviously, this sets the number of tasks used to 2.

You may also want to do this to make the input and output files be in the local file system rather than in the Hadoop
distributed file system (HDFS):

```
conf.set("fs.default.name", "local");
```

You can also set these configuration parameters in `hama-site.xml`. The configuration files should appear somewhere in your program's
class path when the program runs.

2. Run the small input on a 1 node cluster. This will smoke out all of the issues that happen with
distribution and the "real" task runner, but you only have a single place to look at logs. Besides the task logs, the most
useful ones are the grooms and bspmaster logs. Make sure you are logging at the INFO level or you will
miss clues like the output of your tasks.

Use of Log4J in BSP Applications

First of all, you should import the classes of Log4J client API by adding the following import statements at the beginning of your BSP application.

```
import org.apache.commons.logging.Log;
import org.apache.commons.logging.LogFactory;
```

The below example logs INFO level messages by adding line: `LOG.info(peer.getPeerName() + ": Logging test: " + data);` within bsp()
method of PiEstimator example.

```
  public static class MyEstimator extends
      BSP<NullWritable, NullWritable, Text, DoubleWritable, DoubleWritable> {

    ...
    public static final Log LOG = LogFactory.getLog(MyEstimator.class);
    ...

    @Override
    public void bsp(
        BSPPeer<NullWritable, NullWritable, Text, DoubleWritable, DoubleWritable> peer)
        throws IOException, SyncException, InterruptedException {

      int in = 0;
      for (int i = 0; i < iterations; i++) {
        double x = 2.0 * Math.random() - 1.0, y = 2.0 * Math.random() - 1.0;
        if ((Math.sqrt(x * x + y * y) < 1.0)) {
          in++;
        }
      }

      double data = 4.0 * in / iterations;

      LOG.info(peer.getPeerName() + ": Logging test: " + data);
      peer.send(masterTask, new DoubleWritable(data));
      peer.sync();
    }
```

In local mode of Apache Hama, you'll see the INFO messages on console:

```
edward@udanax:~/workspace/hama-trunk$ bin/hama jar examples/target/hama-examples-0.7.0-SNAPSHOT.jar pi
13/05/14 16:02:32 INFO mortbay.log: Logging to org.slf4j.impl.Log4jLoggerAdapter(org.mortbay.log) via org.
mortbay.log.Slf4jLog
13/05/14 16:02:32 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using
builtin-java classes where applicable
13/05/14 16:02:32 INFO bsp.BSPJobClient: Running job: job_localrunner_0001
13/05/14 16:02:32 INFO bsp.LocalBSPRunner: Setting up a new barrier for 10 tasks!
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:6: Logging test: 3.1412
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:9: Logging test: 3.1308
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:5: Logging test: 3.1304
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:4: Logging test: 3.1756
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:3: Logging test: 3.1444
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:8: Logging test: 3.1452
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:0: Logging test: 3.1468
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:1: Logging test: 3.1684
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:2: Logging test: 3.1256
13/05/14 16:02:32 INFO examples.PiEstimator$MyEstimator: local:7: Logging test: 3.114
13/05/14 16:02:35 INFO bsp.BSPJobClient: Current supersteps number: 0
13/05/14 16:02:35 INFO bsp.BSPJobClient: The total number of supersteps: 0
13/05/14 16:02:35 INFO bsp.BSPJobClient: Counters: 7
13/05/14 16:02:35 INFO bsp.BSPJobClient:   org.apache.hama.bsp.JobInProgress$JobCounter
13/05/14 16:02:35 INFO bsp.BSPJobClient:     SUPERSTEPS=0
13/05/14 16:02:35 INFO bsp.BSPJobClient:     LAUNCHED_TASKS=10
13/05/14 16:02:35 INFO bsp.BSPJobClient:   org.apache.hama.bsp.BSPPeerImpl$PeerCounter
13/05/14 16:02:35 INFO bsp.BSPJobClient:     SUPERSTEP_SUM=10
13/05/14 16:02:35 INFO bsp.BSPJobClient:     TIME_IN_SYNC_MS=59
13/05/14 16:02:35 INFO bsp.BSPJobClient:     TOTAL_MESSAGES_SENT=10
13/05/14 16:02:35 INFO bsp.BSPJobClient:     TOTAL_MESSAGES_RECEIVED=10
13/05/14 16:02:35 INFO bsp.BSPJobClient:     TASK_OUTPUT_RECORDS=1
Estimated value of PI is        3.14224
Job Finished in 3.141 seconds
```

In distributed mode of Apache Hama, each BSP task processor creates their own log file under {{$HAMA_HOME}/logs/tasklogs} directory.

```
edward@udanax:~/workspace/hama-trunk$ bin/hama jar examples/target/hama-examples-0.7.0-SNAPSHOT.jar pi
13/05/14 16:14:13 INFO mortbay.log: Logging to org.slf4j.impl.Log4jLoggerAdapter(org.mortbay.log) via org.
mortbay.log.Slf4jLog
13/05/14 16:14:14 INFO bsp.BSPJobClient: Running job: job_201305141614_0001
13/05/14 16:14:17 INFO bsp.BSPJobClient: Current supersteps number: 0
13/05/14 16:14:20 INFO bsp.BSPJobClient: Current supersteps number: 1
13/05/14 16:14:20 INFO bsp.BSPJobClient: The total number of supersteps: 1
13/05/14 16:14:20 INFO bsp.BSPJobClient: Counters: 6
13/05/14 16:14:20 INFO bsp.BSPJobClient:   org.apache.hama.bsp.JobInProgress$JobCounter
13/05/14 16:14:20 INFO bsp.BSPJobClient:     SUPERSTEPS=1
13/05/14 16:14:20 INFO bsp.BSPJobClient:     LAUNCHED_TASKS=3
13/05/14 16:14:20 INFO bsp.BSPJobClient:   org.apache.hama.bsp.BSPPeerImpl$PeerCounter
13/05/14 16:14:20 INFO bsp.BSPJobClient:     SUPERSTEP_SUM=3
13/05/14 16:14:20 INFO bsp.BSPJobClient:     TIME_IN_SYNC_MS=243
13/05/14 16:14:20 INFO bsp.BSPJobClient:     TOTAL_MESSAGES_SENT=3
13/05/14 16:14:20 INFO bsp.BSPJobClient:     TOTAL_MESSAGES_RECEIVED=3
Estimated value of PI is        3.1460000000000004
Job Finished in 6.396 seconds
edward@udanax:~/workspace/hama-trunk$ cat logs/tasklogs/job_201305141614_0001/attempt_201305141614_0001_00000
attempt_201305141614_0001_000000_0.err  attempt_201305141614_0001_000001_0.err
attempt_201305141614_0001_000002_0.err
attempt_201305141614_0001_000000_0.log  attempt_201305141614_0001_000001_0.log
attempt_201305141614_0001_000002_0.log
edward@udanax:~/workspace/hama-trunk$ cat logs/tasklogs/job_201305141614_0001
/attempt_201305141614_0001_000000_0.log
...
13/05/14 16:14:16 INFO examples.PiEstimator$MyEstimator: localhost:61003Logging test: 3.1496
...
```