

Workflow Language (xWFL2.0)

1. Structure

In order to allow users to describe tasks and their dependencies, we defined a simple and flexible XML-based workflow language (xWFL). The workflow language provides the means to build new applications by linking standalone applications.

xWFL2.0 extends xWFL1.0 to support parameter sweep tasks and multiple middleware offered by the Gridbus broker. Our workflow description expressed in XML format, and its structure is shown in Figure 1. It consists of three parts, namely parameter definitions, task definitions and data link definitions.

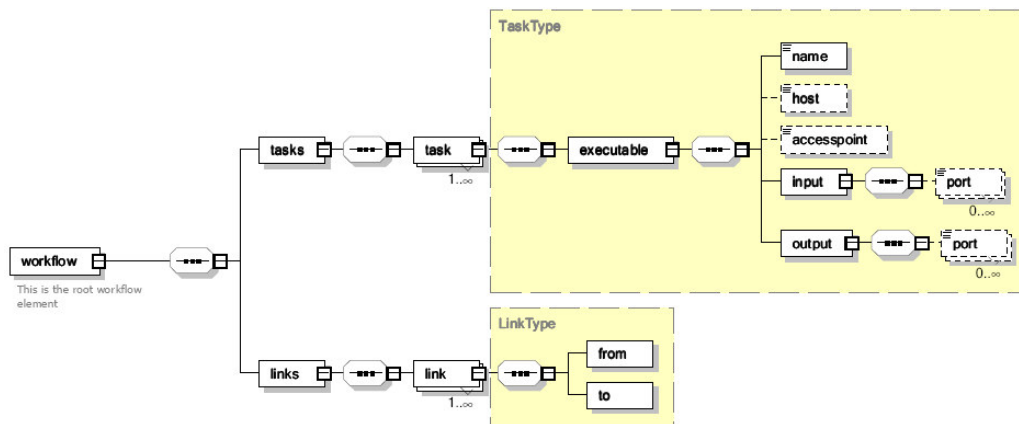


Figure 1. The Structure of Workflow Language.

Task Definition

Basically, `<tasks>` is a set of tasks that are to be executed. Our workflow language supports allow users both abstract workflow [1] and concrete workflow [1]. As shown in the Figure 1, `host` and `accesspoint` are optional. That is the user or higher workflow planner, can either specify the location of a particular resource providing required application services or leave it to the engine to identify their providers dynamically at run-time. In the below example, task A executes `dock.exe` program on host `services.gridbus.com` in the directory `/services`. The executable `dock` has two input I/O ports: port0 (a file) and port1 (a parameter value). The example shows task A only has one output named port2.

```
<tasks>
<task name= "A" >
  <executable>
    <name>dock</name>
    <host >services.gridbus.com</host>
    <accesspoint type= "GlobusGram">/dock.exe</accesspoint>
    <input>
      <port number = "0" type= "file" url= "http://www.gridbus.org/
        datacenter/dock.in">dock.in</port>
      <port number = "1" type= "msg">1</port>
    </input>
    <output>
```

```

    <port number = "2" type = "file">dock.out</port>
  </output>
</executable>
.....
</task>
<task name="B" useBroker="True">
  <executable>
    <name>mycalc</name>
    <host ssh="false" resType="globus">belle.cs.mu.oz.au</host>
    <accesspoint>/home/bellegrid/workflowengine/mycalc/mycalc.sh</accesspoint>
    <input>
      <port number = "0" type="file">para</port>
      <port number = "1" type="msg">0.4</port>
    </input>
    <output>
      <port number = "2" type="file">output</port>
    </output>
  </executable>
</task>
.....
</tasks>

```

The difference between task A and task B in the example is that task B uses the Gridbus broker which is able to support more middleware including globus 2.4, globus 4, pbs, sge and linux fork. Details please refer to the gridbus broker manual on www.gridbus.org/broker. For a task using broker, users need to assign a value to *resType* in *<host>* element. The resource type translation is shown in Table I.

Table I. The resource type translation table.

"pbs"	Portable Batch System
"fork"	Linux fork
"sge"	Sun Grid Engine
"globus"	Globus 2.x
"globus4"	Globus 4

2. Data Links

Data link is used to specify the data flow of the tasks. Below is the data flow description of the workflow in Figure 3. The inputs of task B and task C rely on the output of A. The task A's output needs to be transferred to the node on which task B and task C are executed. Input could be a file, parameter value or data stream.

```

<workflow>
  <tasks>
    <task name="A">
      .....
    </task>
    <task name="B">
      .....
    </task>
    <task name="C">
      .....
    </task>
    <task name="D">
      .....
    </task>
  </tasks>
  <links>
    <link>
      <from task="A" port="2" />
      <to task="B" port="0" />
    </link>
    <link>
      <from task="A" port="2" />
      <to task="C" port="0" />
    </link>
    <link>
      <from task="B" port="1" />
      <to task="D" port="0" />
    </link>
  </links>

```

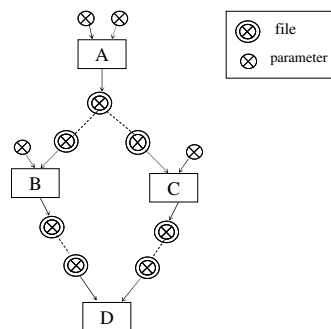


Figure 2. Flow Diagram of Task A, B, C and D.

```
<link>
  <from task="C" port="2" />
  <to task="D" port="1" />
</link>
</links>
</workflow>
```

3. Example

Entire workflow description for Figure 2 is presented in Figure 3. In this example, task A and task B are specified to submit to manjra.cs.mu.oz.au and belle.cs.mu.oz.au respectively. The location of servers to run the rest of the tasks will be discovered by workflow engine at the run-time via the Grid Market Directory (www.gridbus.org/gmd).

```

<workflow>
  <tasks>
    <task name="A" useBroker="true">
      <executable>
        <name>calc</name>
        <host ssh="true" resType="pbs" user="test" pass="testworkflow" queue="workq">manjra.cs.mu.oz.au</host>
        <accesspoint>/home/globus/calc</accesspoint>
        <input>
          <port number="0" type="msg">3</port>
          <port number="1" type="msg">0.1</port>
        </input>
        <output>
          <port number="2" type="file">output</port>
        </output>
      </executable>
    </task>
    <task name="B" useBroker="True">
      <executable>
        <name>mycalc</name>
        <host ssh="false" resType="globus">belle.cs.mu.oz.au</host>
        <accesspoint>/home/bellegrid/workflowengine/mycalc/mycalc.sh</accesspoint>
        <input>
          <port number="0" type="file">para</port>
          <port number="1" type="msg">0.4</port>
        </input>
        <output>
          <port number="2" type="file">output</port>
        </output>
      </executable>
    </task>
    <task name="C">
      <executable>
        <name>mycalc</name>
        <host>belle.cs.mu.oz.au</host>
        <accesspoint type="GT2Gram">/home/bellegrid/workflowengine/mycalc/mycalc.sh</accesspoint>
        <input>
          <port number="0" type="file">para</port>
          <port number="1" type="msg">0.5</port>
        </input>
        <output>
          <port number="2" type="file">output</port>
        </output>
      </executable>
    </task>
    <task name="D">
      <executable>
        <name>addcalc</name>
        <input>
          <port number="0" type="file">para1</port>
          <port number="1" type="file">para2</port>
        </input>
        <output>
          <port number="2" type="file">output</port>
        </output>
      </executable>
    </task>
  </tasks>
  <links>
    <link>
      <from task="A" port="2" />
      <to task="B" port="0" />
    </link>
    <link>
      <from task="A" port="2" />
      <to task="C" port="0" />
    </link>
    <link>
      <from task="B" port="2" </from>
      <to task="D" port="0" </to>
    </link>
    <link>
      <from task="C" port="2" </from>
      <to task="D" port="1" </to>
    </link>
  </links>
</workflow>

```

Figure 3. Workflow Diagram of the Experiment.

References:

[1] J. Yu and R. Buyya, A Taxonomy of Workflow Management Systems for Grid Computing, Journal of Grid Computing, Volume 3, Numbers 3-4, Pages: 171-200, Springer Science+Business Media B.V., New York, USA, Sept. 2005.