July 13, 2004

Agenda

JSDL XML/XSD

Attendees

Ali A.
Andreas S.
Dan T.
Steve M.
Sayaka Akioka
Anuj M.
Kazushige Saga

Actions
---
o Andreas to write a brief how-to and intro to the JSDL tracking system on GridForge.
o Steve to drop a e-mail to the list to propose a solution for chicken and egg problem of Host(Name) and Resource, and the level of specificity that we want for each, i.e. do we want to be able to specify particular Env Vars for each Host(Name) and/or resource as well as for the entire JSDL request.
o Ali to send request to the JSDL list to ask people to try out the latest schema.
o Andreas to send an example of a three-tier system job request to the list.
o One of us to chase up Chris with a possible solution to the ProcessTopolgy problem in the schema.
o Steve to send an example of use of the inheritence capability to the JSDL list and to sound out comments.
o What do we do with logical operators, explicit, if so who shall we copy?

[Text Slide A]

```
<?xml version="1.0" encoding="UTF-8"?>
<jsdl:job xmlns:jsdl="http://www.gridforum.org/JSDL"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="
        http://www.gridforum.org/JSDL
        jsdl.xsd">
<!--
================================================================================-
->
<!-- The first section of the JSDL document describes the general features of -->
<!-- a job. In the common case this section is all that you will need.       -->
<!--
================================================================================-
->

  <jsdl:JobIdentification>
    <jsdl:JobName>MyFirstJob</jsdl:JobName>
    <jsdl:JobAnnotation>This is the first job ever submitted through JSDL</jsdl:JobAnnotation>
    <jsdl:JobAnnotation>This job won't really do anything</jsdl:JobAnnotation>

    <jsdl:ExecutionUserID>
      <jsdl:User name="steve" group="myGroup" />
    </jsdl:ExecutionUserID>
    <jsdl:JobProjectName>WritingTheSpecProject</jsdl:JobProjectName>
    <jsdl:JobCategory>writingJob</jsdl:JobCategory>
    <jsdl:Extend name="PBSQueueName">myQueue</jsdl:Extend>
  </jsdl:JobIdentification>
```

[Text Slide B]

```
 <jsdl:Resource name="writingPerson">
   <!-- The following describes the general architecture that should be used -->
   <!-- for all resources in the job. Ie all boxes will fall within these    -->
   <!-- requirements.                                                        -->
   <!-- Unless the job is complex (eg MPI) then this is all that is needed   -->

   <!-- Only use intel and spark processors for this job                     -->
   <jsdl:CPUDescription>itel8086</jsdl:CPUDescription>

   <jsdl:CPUDescription>spark</jsdl:CPUDescription>
   <!-- Each box used must have at least 1 processor                         -->
   <jsdl:CPUCount>1</jsdl:CPUCount>
   <!-- All boxes must run at at least 8Mhz                                  -->
   <jsdl:CPUSpeed>8Mhz</jsdl:CPUSpeed>
   <!-- Each box must have at least 1024 Mb of memory                        -->
   <jsdl:PhysicalMemory>1024Mb</jsdl:PhysicalMemory>

   <!-- See usage record for units                                           -->
   <!-- **## Move Software Requirements and Environment Variables Here ?? -->
 </jsdl:Resource>

 <jsdl:Resource name="editor" derivedFrom="writingPerson">
   <!-- The "editor" takes on all values defined in "writingPerson" with the -->
   <!-- following additions                                                  -->
   <jsdl:TemporarySpace>10Gb</jsdl:TemporarySpace>
 </jsdl:Resource>

 <!-- Process topology - we need to define this some people may like to      -->
 <!-- try writing some examples based on their needs here.                   -->
 <jsdl:ProcessTopology>?????</jsdl:ProcessTopology>

 <!-- Set up environmental variables                                         -->
<jsdl:Environment name="baseEnv">
   <!-- These variables apply to all boxes used                              -->
  <jsdl:EnvironmentVariable name="TASKID">Some Value</jsdl:EnvironmentVariable>

  <jsdl:EnvironmentVariable name="PROBLEMSIZE">1000</jsdl:EnvironmentVariable>
</jsdl:Environment>

 <jsdl:Environment name="compelxEnv" derivedFrom="baseEnv">
```

```
  <!-- Takes all environment elements from "baseEnv" and adds the following -->
  <jsdl:EnvironmentVariable name="OFFSET">-1</jsdl:EnvironmentVariable>
 </jsdl:Environment>

 <jsdl:SoftwareRequirements name="simpleSoft">

  <jsdl:OperatingSystem>
   <jsdl:OperatingSystemDescription>Windows2000</jsdl:OperatingSystemDescription>
   <jsdl:OperatingSystemVersion>2000</jsdl:OperatingSystemVersion>
  </jsdl:OperatingSystem>
  <jsdl:OperatingSystem>
   <jsdl:OperatingSystemDescription>Linux</jsdl:OperatingSystemDescription>
   <jsdl:OperatingSystemVersion>RedHat7.2</jsdl:OperatingSystemVersion>

  </jsdl:OperatingSystem>
  <jsdl:Limits>
   <jsdl:ProcessVirtualMemoryLimit>1Gb</jsdl:ProcessVirtualMemoryLimit>
   <jsdl:VirtualMemoryLimit>1Gb</jsdl:VirtualMemoryLimit>
   <jsdl:DataSegmentSizeLimit>10Gb</jsdl:DataSegmentSizeLimit>
   <jsdl:CoreDumpSizeLimit>0</jsdl:CoreDumpSizeLimit>
   <jsdl:CPUTimeLimit>10Hours</jsdl:CPUTimeLimit>

   <jsdl:WallTimeLimit>24Hours</jsdl:WallTimeLimit>
  </jsdl:Limits>
  <jsdl:Queue>myQueue</jsdl:Queue>
  <jsdl:Queue>anotherQueue</jsdl:Queue>
 </jsdl:SoftwareRequirements>

 <jsdl:Application>
  <jsdl:ExecutableDescription>Run the vi command</jsdl:ExecutableDescription>

  <jsdl:ExecutableName type="bash">vi</jsdl:ExecutableName>
  <jsdl:Argument>-noX</jsdl:Argument>
  <jsdl:Argument>myFile.txt</jsdl:Argument>
  <jsdl:StdIn>myin.txt</jsdl:StdIn>
  <jsdl:StdOut>myOut.txt</jsdl:StdOut>
  <jsdl:StdErr>myErr.txt</jsdl:StdErr>

  <jsdl:WorkingDirectory>myBrainDump</jsdl:WorkingDirectory>
  <jsdl:Log>processLog.txt</jsdl:Log>
```

```
  </jsdl:Application>

<!--
=============================================================================-
->
<!-- In this section of the document we describe a single host that can be    -->
<!-- used to run the job on. This section defines a resource either           -->
<!-- explicitly named or defined by the characteristics defined within        -->
<!-- note that the following should be seen as extending the definition of a  -->
<!-- resource specified above in the general section.                         -->
<!--
=============================================================================-
->
  <jsdl:Host name="bar" description="This is a specific host type">

    <!-- List of host names we know match the requirements                   -->
    <jsdl:HostName>myFastBox.ggf.org</jsdl:HostName>
    <jsdl:HostName>anotherGoodBox.ggf.org</jsdl:HostName>
    <!-- Valid user id's that can be used on these boxes                     -->
    <jsdl:ExecutionUserID>
      <jsdl:User name="asm" group="secondGroup" />
    </jsdl:ExecutionUserID>
    <jsdl:Resource>

      <!-- The following describes the general architecture that should be    -->
      <!-- used for this resources in the job. Ie boxes matching this         -->
      <!-- description can be used for this host "id".                        -->

      <!-- Only use intel and spark processors for this job                  -->
      <jsdl:CPUDescription>itel8486</jsdl:CPUDescription>
      <jsdl:CPUDescription>spark2</jsdl:CPUDescription>
      <!-- Each box used must have at least 2 processors                     -->
      <jsdl:CPUCount>2</jsdl:CPUCount>

      <!-- All boxes must run at at least 1024Mhz                            -->
      <jsdl:CPUSpeed>1024Mhz</jsdl:CPUSpeed>
      <!-- Each box must have at least 4096 Mb of memory                     -->
      <jsdl:PhysicalMemory>4096Mb</jsdl:PhysicalMemory>
      <!-- See usage record for units                                        -->
    </jsdl:Resource>
```

```
 <!-- Process topology - we need to define this some people may like to    -->

 <!-- try writing some examples based on their needs here.                 -->
 <jsdl:ProcessTopology>?????</jsdl:ProcessTopology>

 <!-- Set up environmental variables                                       -->
<jsdl:Environment>
  <!-- These variables apply to all boxes used                            -->
 <jsdl:EnvironmentVariable name="TASKID">Some new Value</jsdl:EnvironmentVariable>
 <jsdl:EnvironmentVariable name="PROBLEMSIZE">2000</jsdl:EnvironmentVariable>

 <jsdl:EnvironmentVariable name="CORRECTION">-1</jsdl:EnvironmentVariable>
</jsdl:Environment>

 <jsdl:SoftwareRequirements>
  <jsdl:OperatingSystem>
   <jsdl:OperatingSystemDescription>Windows2000</jsdl:OperatingSystemDescription>
   <jsdl:OperatingSystemVersion>2000.0.1.3</jsdl:OperatingSystemVersion>
  </jsdl:OperatingSystem>
  <jsdl:OperatingSystem>

   <jsdl:OperatingSystemDescription>Linux</jsdl:OperatingSystemDescription>
   <jsdl:OperatingSystemVersion>RedHat7.2</jsdl:OperatingSystemVersion>
   <jsdl:Extend name="patchVersion">October2003</jsdl:Extend>
  </jsdl:OperatingSystem>
  <jsdl:Limits>
   <jsdl:ProcessVirtualMemoryLimit>2Gb</jsdl:ProcessVirtualMemoryLimit>
   <jsdl:VirtualMemoryLimit>0.5Gb</jsdl:VirtualMemoryLimit>

   <jsdl:DataSegmentSizeLimit>15Gb</jsdl:DataSegmentSizeLimit>
   <jsdl:CoreDumpSizeLimit>2bytes</jsdl:CoreDumpSizeLimit>
   <jsdl:CPUTimeLimit>11Hours</jsdl:CPUTimeLimit>
   <jsdl:WallTimeLimit>48Hours</jsdl:WallTimeLimit>
  </jsdl:Limits>
  <jsdl:Queue>brokenQueue</jsdl:Queue>

  <jsdl:Queue>workingQueue</jsdl:Queue>
 </jsdl:SoftwareRequirements>
```

```
    <!-- Allow a process topology to be specified here                    -->
  </jsdl:Host>


<!--
=========================================================================-
->
<!-- In this section of the document we describe a host grouping that can be  -->
<!-- used to run the job on. This group defines a collection of resources     -->
<!-- either explicitaly named or defined by the characteristics defined within-->
<!-- note that the following should be seen as extending the definition of a  -->
<!-- resource specified above in the general section.                         -->

<!--
=========================================================================-
->
  <jsdl:HostGroup name="foo" description="This host type can be used for our job">

    <!-- Some examples of the resources that match this host type. We do not  -->
    <!-- limit the system to only use these resources - but they are          -->
    <!-- probably a good starting point.                                      -->
    <jsdl:HostName>mybox.ggf.org</jsdl:HostName>
    <jsdl:HostName>secondbox.ggf.org</jsdl:HostName>
    <jsdl:HostName>resourcePool1</jsdl:HostName>

    <!-- Valid user id's that can be used on these boxes                      -->
    <jsdl:ExecutionUserID>
      <jsdl:User name="asm2" group="thirdGroup" />
    </jsdl:ExecutionUserID>

    <jsdl:Resource>
      <!-- Hosts in this type are more narrowly specified                   -->
      <!-- Only use intel and spark processors for this job                 -->
      <jsdl:CPUDescription>itelPentium4</jsdl:CPUDescription>

      <!-- Each box used must have at least 4 processors                      -->
      <jsdl:CPUCount>4</jsdl:CPUCount>
      <!-- All boxes must run at at least 2048Mhz                           -->
      <jsdl:CPUSpeed>2048Mhz</jsdl:CPUSpeed>
      <!-- Each box must have at least 4096 Mb of memory                      -->
```

```
      <jsdl:PhysicalMemory>4096Mb</jsdl:PhysicalMemory>
      <!-- See usage record for units                                    -->

     <jsdl:Network>
       <!-- Boxes of this type must have fast ethernet cards             -->
      <jsdl:NetworkDescription>FastEthernet</jsdl:NetworkDescription>
       <!-- and at least two cards per box                               -->
      <jsdl:NetworkCount>2</jsdl:NetworkCount>
       <!-- Capable of supporting 100MB transfers                        -->
      <jsdl:NetworkBandwidth>100MB</jsdl:NetworkBandwidth>

     </jsdl:Network>
      <!-- Provide a file space called scratch which can store at least  -->
      <!-- 4096Mb of data                                                -->
      <jsdl:FileSystem name="scratch" size="4096Mb"/>
      <!-- Have (or be able to have" the following filesystem mounted    -->
     <jsdl:FileSystem name="myBrainDump"
size="10GB">nfs:hostname:/export/vol01</jsdl:FileSystem>
      <!-- Have 100GB of swap space                                      -->
     <jsdl:SwapSpace>100GB</jsdl:SwapSpace>

      <!-- Have temp space of 200KB of data                              -->
     <jsdl:TemporarySpace>200KB</jsdl:TemporarySpace>
      <!-- Our job requires exclusive use of these nodes?               -->
     <jsdl:ExclusiveExecution>true</jsdl:ExclusiveExecution>
    </jsdl:Resource>

    <!-- Process topology - we need to define this some people may like to  -->
    <!-- try writing some examples based on their needs here.               -->

    <jsdl:ProcessTopology>?????</jsdl:ProcessTopology>

    <!-- Set up environmental variables                                -->
   <jsdl:Environment>
     <!-- These variables apply to all boxes used                      -->
    <jsdl:EnvironmentVariable name="CORRECTION">-2</jsdl:EnvironmentVariable>
   </jsdl:Environment>

    <jsdl:SoftwareRequirements>
      <jsdl:Limits>
```

```
        <jsdl:ProcessVirtualMemoryLimit>4Gb</jsdl:ProcessVirtualMemoryLimit>

        <jsdl:VirtualMemoryLimit>0.5Gb</jsdl:VirtualMemoryLimit>
        <jsdl:DataSegmentSizeLimit>75Gb</jsdl:DataSegmentSizeLimit>
      </jsdl:Limits>
      <jsdl:Queue>brokenQueue</jsdl:Queue>
      <jsdl:Queue>workingQueue</jsdl:Queue>
    </jsdl:SoftwareRequirements>
  </jsdl:HostGroup>

<!--
================================================================================-
->
<!-- In this section the files that are required to be staged in/out of the   -->
<!-- resource are defined and the possible locations where they may be staged -->
<!-- to/from can be defined.                                                  -->
<!--
================================================================================-
->

 <jsdl:DataStaging> <!-- **## Update this in the xsd -->
    <jsdl:File>
     <jsdl:FileName>myFirstFile.txt</jsdl:FileName>
     <jsdl:FileSystemName>myBrainDump</jsdl:FileSystemName>

      <jsdl:Source>http://www.ggf.org/myFirstFile.txt</jsdl:Source>
      <jsdl:CreationFlag>overwrite</jsdl:CreationFlag>
    </jsdl:File>
    <jsdl:File>
      <jsdl:FileName>mySecondFile.data</jsdl:FileName>
     <jsdl:FileSystemName>myBrainDump</jsdl:FileSystemName>
      <jsdl:Target>gsiftp::box.ggf.org/homes/ggf/mySecondFile.data</jsdl:Target>

      <jsdl:CreationFlag>append</jsdl:CreationFlag>
    </jsdl:File>
    <jsdl:File>
      <jsdl:FileName>StageThrough.data</jsdl:FileName>
     <jsdl:FileSystemName>scratch</jsdl:FileSystemName>
      <jsdl:Source>ftp://ftp.ggf.org/largeData/input.dat</jsdl:Source>
      <jsdl:Source>ftp://ftp.slowbox.org/largeDataBackup/record.data</jsdl:Source>
```

```
      <jsdl:Target>gsiftp::box.ggf.org/homes/ggf/result.data</jsdl:Target>
      <jsdl:CreationFlag>overwrite</jsdl:CreationFlag>
    </jsdl:File>
 </jsdl:DataStaging>
</jsdl:job>
```

[Text Slide C]

```
<!-- Default Overall ------------------------------------------------------------------------ -->
<ExecutionUserID>
</ExecutionUserID>
<OperatingSystem>
</OperatingSystem>
<ProcessTopology>
</ProcessTopology>
<?>
</?>
<Resource>
<!-- Default general resource requirements for the job... -->
</Resource>
<Environment>
<!-- Default general env requirements for the job... -->
</Environment>
<Limits>
<!-- Default general limits for the job... -->
</Limits>

<!-- Specific Host -------------------------------------------------------------------------- -->
<Host>
<HostName>
<!-- HostName for this specific Host -->
</HostName>
<ExecutionUserID>
</ExecutionUserID>
<OperatingSystem>
</OperatingSystem>
<ProcessTopology>
</ProcessTopology>
<Resource>
<!-- Resource requirements specific to this Host... -->
</Resource>
<Environment>
<!-- Env requirements specific to this Host... -->
</Environment>
<Limits>
<!-- Limits requirements specific to this Host... -->
</Limits>
```

```
</Host>

<!-- Host Group ---------------------------------------------------------------------- -->
<HostGroup>
<ExecutionUserID>
</ExecutionUserID>
<OperatingSystem>
</OperatingSystem>
<ProcessTopology>
</ProcessTopology>
<Resource>
<!-- Resource requirements specific to the HostGroup... -->
</Resource>
<Environment>
<!-- Env requirements specific to the HostGroup... -->
</Environment>
<Limits>
<!-- Limits requirements specific to this HostGroup... -->
</Limits>
<!-- Host ---------------------------- -->
<Host>
<HostName>
<!-- HostName for this specific Host -->
</HostName>
<ExecutionUserID>
</ExecutionUserID>
<OperatingSystem>
</OperatingSystem>
<ProcessTopology>
</ProcessTopology>
<Resource>
<!-- Resource requirements specific to this Host... -->
</Resource>
<Environment>
<!-- Env requirements specific to this Host... -->
</Environment>
<Limits>
<!-- Limits requirements specific to this Host... -->
</Limits>
</Host>
```

<!-- Host ---------------------------- -->
...
<!-- Host ---------------------------- -->
...
</HostGroup>
Untitled
<!-- -- -->

[Text Slide E]

```
<CPU>
  <OR>
    <type1>intel</type1>
    <type2>spqark</type2>
  </OR>
<CPU>

<OR>
  <CPU>Spark <ProcessorSpeed>100</ProcessorSpeed></CPU>
  <CPU>intel</CPU>
</OR>

<OR>
  <ProcessorSpeed>100</ProcessorSpeed>
  <ProcessorSpeed>200</ProcessorSpeed>
</OR>

<ALL>

</ALL>

<EXACTLYONE>

</EXACTLYONE>
```

[Text Slide F]