Leveraging the standardized JSDL Paramater Sweep Extension to enhance Scientific Workflow Execution on High Performance Computing resources

Thursday, 11 April 2013 16:00 (30 minutes)

Impact

Our Taverna based bioinformatics use case showed, how parameter sweeps benefit researchers in saving client resources and runtime. Certainly, the usability of the UNICORE parameter sweep implementation and the parameter sweep library itself is not only limited to the Taverna Workflow management System and UNICORE. As the library implements an abstract application programming interface for the handling of a standardized JSDL sweep job, it can be included and implemented by various middleware systems. In doing so, many distributed computing infrastructures (DCI) can enhance their handling of sweeps or job bunches. As parameter or file sweeps are a typical scenario for various scientific disciplines an improved and standardized handling of job sweeps results in a widespread and significant improvement of runtime and thus saving of computing resources. Furthermore, scientific communities can easily adopt their workflow clients to the available parameter sweep mechanism. This client side adoption is feasible, as the sweep mechanism itself is not focused to a specific kind of research area or application. For example in fusion science or drug discovery, workflow management systems are used widely in combination with distributed environments. In these cases, our standards based parameter sweep extension could support other scientific workflow management systems. Overall we showed that the parameter sweep approach is a very promising mechanism to reduce client side workload as well as usage of compute, data, and network resources.

Summary

Our talk will focus on how to make the execution of scientific workflows more effective and highly performant by using the standardized JSDL parameter sweep extension. As a scientific use case we have chosen an application from the bioinformatics domain known as LOCUSTRA workflow. This workflow helps in predicting the secondary structure of a protein from its amino acid sequence by using support vector machine classification. The given use case is realized via UNICORE's parameter sweep implementation as an interface for managing and incarnating parametric jobs on a local resource management system. On the client side we have developed an extension plugin to the Taverna workflow management system, which allow users to create multiple workflow components with implicit iteration represented by a single JSDL parameter sweep instance.

Description

Certain scientific use cases possess complex requirements to have Grid jobs executed in collections where jobs' request contains only some variation in different parts. These scenarios can easily be tackled by a single job request which abstract this variation and can represent the same collection. We can find several examples in the domain of high energy physics, biology, and chemistry etc, where parametric jobs are handled manually i.e. creating individual job requests. OGF standards community model this requirement through the JSDL Parameter Sweep specification which takes a modular approach to handle different type of parameter sweeps, such as Document and File Sweep. UNICORE server environment implements this specification build upon its existing JSDL implementation. The actual realization is carried out by the UNICORE's execution management system known as XNJS. In the case of sweep jobs XNJS detects the job request and create sweeps if it contains the JSDL sweep elements. After detection phase the jobs are incarnated on a target resource management system. In our talk we will unleash more technical details on sweep oriented collection of jobs. From an application perspective we have taken LOCUSTRA bioinformatics workflow application. Initially this application was manually triggering 120 jobs part of the same use case. As this collection has only some variation in the

application arguments, therefore we took this example and modelled it using JSDL parameter sweep. It really helped our use case as it was evident in terms of improved job execution and less monitoring overhead, both from the server and client perspective. As far as the user facing end is concerned the application is currently being submitted via Taverna Workflow management system. Therein we extended Taverna in the form of a plugin responsible for constructing job request payloads compliant with the JSDL parameter sweep schema, and then dispatch it to the UNICORE web services front end.

Primary author: MEMON, Shahbaz (JUELICH)

Co-authors: Mr DEMUTH, Bastian (Forschungszentrum Juelich GmbH); Mr RIEDEL, MORRIS (JUELICH SUPERCOMPUTING CENTRE); Ms HOLL, Sonja (Forschungszentrum Juelich GmbH)

Presenter: MEMON, Shahbaz (JUELICH)

Session Classification: Workflow Collaboration Workshop