Enabling Distributed Applications with SAGA

João Abecasis, Shantenu Jha, Hartmut Kaiser, Joohyun Kim, André Merzky, and Ole Weidner

Center for Computation & Technology, Louisiana State University, Baton Rouge, U.S.A.

LSU — CENTER FOR COMPUTATION & TECHNOLOGY

NSF

Abstract

The Simple API for Grid Applications (SAGA), a proposed recommendation of the Open Grid Forum (OGF), defines a high-level programmatic interface for developers of Distributed Applications [1]. The fundamental idea of SAGA is to lower the barrier for applications and application scientists to utilize distributed infrastructure. SAGA provides a simple, uniform, stable interface to the most often required functionality in order to construct general purpose, extensible and scalable applications.

Our group has lead the SAGA effort, starting from the specification effort at the OGF to providing the first C++ implementation [2]. We are also developing several different novel applications, using SAGA to harness the power of distributed infrastructure.

SAGA has already been used to develop different types of distributed applications. Namely, (i) converting legacy applications to utilize distributed resources; (ii) development of applications based upon abstractions and frameworks that are themselves developed using SAGA; (iii) first principles applications, explicitly cognizant of the fact that they will operate in a distributed environment, where the application logic is coupled with the distributed logic. SAGA supports the development of these applications and many others, thus providing a tool to develop a broad and general class of applications.
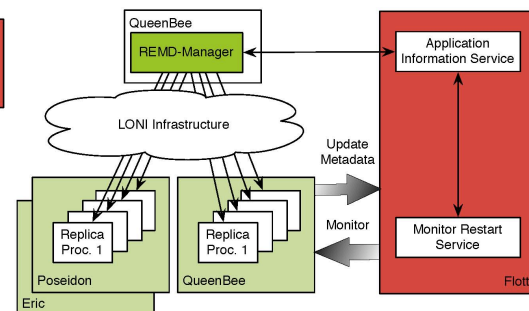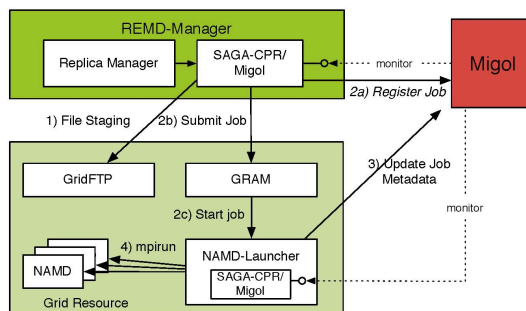
Simple, Powerful Abstraction Layer

SAGA facilitates the use of distributed infrastructure by providing a simple interface across different middleware distributions and environments. Therefore once an application has been written using SAGA it can be deployed and run on any environment in which SAGA is supported.

We are developing adaptors for the most commonly occurring distributed environments. Additionally SAGA provides the abstractions from which commonly occurring execution patterns and usage modes can be supported. For example for data-intensive applications, we create a framework that supports the common MapReduce pattern. Applications involving basic functionality such as searching, can then be deployed over distributed environments

Connections with CyberTools

SAGA is being used within the Cybertools project in several critical ways:

o It is being used to create a general purpose "Application Manager", that will enable many science drivers to utilize remote LONI machines without any changes to the execution environment. In particular it can be used to support specific application usage patterns, for example, it has been used for distributed replica-exchange (RE) simulations using NAMD. The same infrastructure can be used for use with other codes such as LAMMPS, etc. The figure above provides details on how SAGA is used to implement RE.

o SAGA will be the interfaced with Cactus applications to use Information Services and other advanced CyberInfrastructure features.

o SAGA will also provide the basic capability for interfacing multi-physics applications (via extension to the API to support messaging)

References

1. Goodale, T, Jha, S, Kaiser, H, Kielmann, T, Kleijer, P, Merzky, A, Shalf, J, Smith, C, (2007) GFD-R-P.90 A Simple API for Grid Applications (SAGA), Open Grid Forum
2. SAGA C++ Project [Online]. http://saga.cct.lsu.edu