Extending Rock Physics to the Cloud and Beyond

[Poster Abstract]

Ashley Zebrowski
Rutgers University, Electrical & Computer Engineering
94 Brett Road Piscataway, NJ 08854-8058
anz7@rutgers.edu

Rosa Filgueira
University of Edinburgh, School of Informatics
10 Crichton Street
Edinburgh, EH8 9AB
rosa.filgueira@ed.ac.uk

Shantenu Jha
Rutgers University, Electrical & Computer Engineering
94 Brett Road Piscataway, NJ 08854-8058
shantenu.jha@rutgers.edu

Malcolm Atkinson
University of Edinburgh, School of Informatics
10 Crichton Street
Edinburgh, EH8 9AB
mpa@staffmail.ed.ac.uk

1. INTRODUCTION

Extending a single-resource application outwards onto distributed resources presents a multitude of challenges from a software engineering perspective (e.g. appropriate choice of abstractions and algorithms) in addition to fundamental questions regarding the location and placement of data. The impact of these challenges is further amplified when applications are accessed by multiple simultaneous users, each executing their own models, presenting opportunities to experiment at-scale with regards to distributing work and the total amount of scientific data which is sourced and produced.

Supported by the OSDC-PIRE[1] program which provides international research and education experience for students, a single-machine multi-user rock physics forecasting engine is extended to make use of cloud resources (OSDC[2]) as well as HPC (XSEDE[3]) resources. Through a web-accessible gateway, multiple rock physics scientists may log on and submit their own workflows which execute on the requested data onto multiple supported machines. In addition to handling management of jobs and data to distributed resources, the gateway also provides visualization functions to aid understanding of the executing forecast models.

2. BACKGROUND AND RELATED WORK

The two main drivers for this work are the rock physics simulation and gateway software EFFORT[4] [1] and the SAGA[5] [2] API for interfacing with distributed computing resources.

2.1 EFFORT

EFFORT is a multi-disciplinary collaboration which aims to determine the predictability of brittle failure rock samples in laboratory experiments. EFFORT aims to explain how this predictability scales with regard to the greater complexity, physical scale, and slower strain-rates of natural world phenomena.

EFFORT consists of several main components. A collection of user-selectable rock physics forecast models enables scientists to set up different forecasting scenarios. EFFORT has a storage system composed by a repository and a database for storing models, experiment data and results updated periodically via FAST[6], which can be accessed as inputs for the forecast models. A web-accessible gateway allows scientists to execute forecast models using data from the database and provide results via a user-friendly interface.

The gateway enables multiple users to make use of EFFORT; users may log in, select particular data, execute rock physics forecast models and visualize the results.

2.2 SAGA

SAGA is an implementation of an OGF[7] technical specification, providing a common consistent high-level API for interacting with distributed computing functionality. SAGA allows application writers to manage computational jobs, data, resources and more via an extensible, adaptor-based interface.

SAGA offers advantages over traditional distributed appli-

[1] Open Science Data Cloud Partnership for International Research and Education
[2] Open Science Data Cloud
[3] eXtreme Science and Engineering Discovery Environment

[4] Earthquake and Failure Forecasting in Real Time Forecasting Model Testing
[5] Simple API for Grid Applications
[6] (Flexible Automated Streaming Transfer)
[7] Open Grid Forum

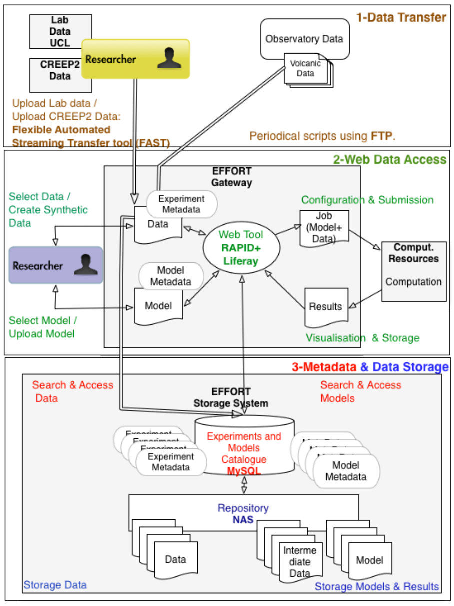

Figure 1: EFFORT Architecture: Architectural diagram illustrating the interaction between users, the web-based EFFORT gateway, and the role of EFFORT's subsystems.

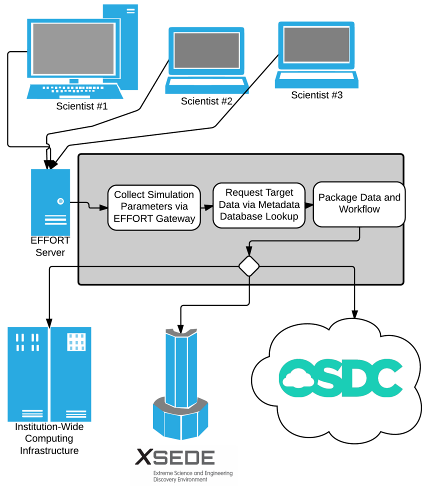

Figure 2: SAGA-EFFORT Architecture: Architectural diagram illustrating the role SAGA plays in distributing jobs to multiple resources.

cation development by abstracting distributed functionality into a single well-defined access layer, allowing multiple distributed middlewares to be targeted via the use of a single API. This is of use in heterogeneous environments, where the cost required to code, debug, and perform upkeep for middleware-specific code may be prohibitive. As this work helps demonstrate, SAGA also lends itself to the construction of higher-level constructs such as workflows, application management systems, runtime environments, and gateways.

3. ACCOMPLISHMENTS

The work has several main goals; first is to demonstrate the use of SAGA in successfully enabling existing applications to make use of distributed computing resources. Work has been performed to understand the movement of data for scientific simulations and to investigate use of the OSDC as a "hub" for scientific storage. Finally, the role of resource assignment in gateway applications has been investigated.

4. RESULTS

Initial results illustrate the validity of the approach, with simulations executed on cloud resources (OSDC) and XSEDE resources (Stampede), in addition to the previously existing local execution capabilities. Full integration with the EFFORT gateway is also underway, to enable more users to make use of the improvements. The investigation of the role of the OSDC data resources to increase the volume of data centralized and made accessible to domain scientists is underway as well.

5. FUTURE WORK

Scheduling and resource assignment of heterogeneous tasks created by scientists on the EFFORT gateway is one avenue for future work, given the increased range of resource utilization made possible by incorporating SAGA with EFFORT.

The usage of databases with distributed data and compute is another potential target for further investigation, as connectivity issues could prove problematic for certain distribution modes.

Increasing the scale of submission with regards to submission rate/core counts/data volume and the number of submission sites could be considered as well.

Another source of future work regards to community uptake – applying these techniques to similar single-user applications could lead to similar gains.

Finally, more investigation could be pursued regarding the management of credentials across distributed resources, particularly tying per-user gateway accounts to the submission to remote resources in a secure, scalable manner.

6. REFERENCES

[1] R. Filgueira, M. Atkinson, A. Bell, I. Main, S. Boon, and P. Meredith. Earthquake and failure forecasting in real-time: A forecasting model testing centre. In *EGU General Assembly Conference Abstracts*, volume 15, page 737, 2013.

[2] T. Goodale, S. Jha, H. Kaiser, T. Kielmann, P. Kleijer, A. Merzky, J. Shalf, and C. Smith. A Simple API for Grid Applications (SAGA). OGF Recommendation Document, GFD.90, Open Grid Forum, 2007. `http://ogf.org/documents/GFD.90.pdf`.