Science Gateway (SG) Approach to Job Execution on Distributed Systems: A Case Study of the Modelling and Simulation Applications for Life Sciences and Healthcare

Adedeji Fabiyi (1), Simon J.E. Taylor (1), Anastasia Anagnostou (1), Mario Torrisi (2), and Roberto Barbera (2)

(1) Modelling and Simulation group, Department of Computer Science, Brunel University London.

(2) Department of Physics and Astronomy of the Catania and INFN via S.Sofia 64, Catania, I-95123, Italy.

Introduction

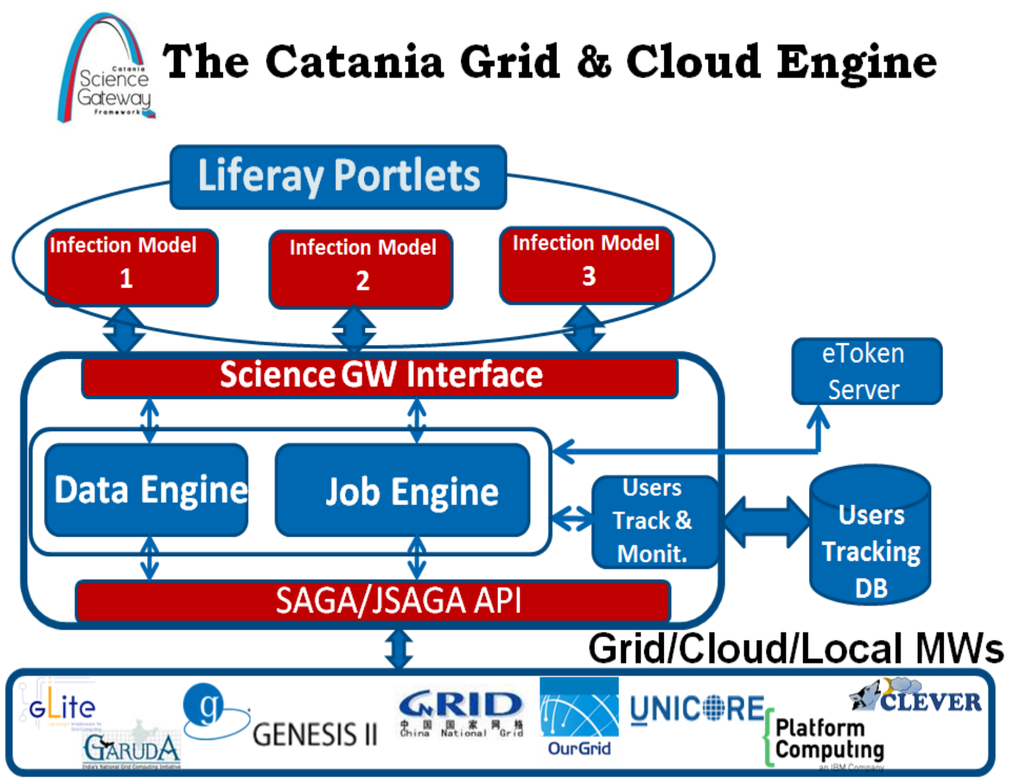
Distributed systems have rapidly evolved from centralised computing/mainframe computing, to cluster computing, in the 80's/90's, to grid computing, in the 90's and 00's, and to cloud computing in the modern era. A consequence of this evolution is that affordable computing resources and powerful machines, with large storage and memory capacity, become readily available. In particular, this could aid the life sciences and health care to execute large modelling and simulation applications that needs massive amount of computational power. However, the deployment and use of these environments can be extremely complex and could be quite a daunting experience which could, in turn, prevent non-ICT experts from adopting these technologies. In order to simplify the use of these environments, otherwise known as e-infrastructures, a science gateway (SG) approach has been proposed. There are different SG frameworks for building and customizing SG instances, for specific scientific domain (Barbera, Fargetta and Rotondo, 2011), (Kocot *et al.*, 2014), (Russell *et al.*, 2008), and (Balasko, Farkas and Kacsuk, 2013), and the commonly used SG frameworks in Europe are seen in the work of the aforementioned authors. However, the Catania Science Gateway Framework (CSGF), that was developed by (Barbera, Fargetta and Rotondo, 2011), has been adopted in this work.

Aims and Objectives

The aim of this research is: to explore new approaches to executing modelling and simulation applications, on worldwide infrastructure, for life sciences and healthcare sector, and consequently develop a science gateway framework for the efficient execution of such applications.

To achieve this aim, the following objectives will be met:

1. Do a thorough literature review and investigate issues surrounding the use of science gateways for the execution of jobs.
2. Develop a Science Gateway (SG) methodology for the execution of Agent Based Simulation (ABS) application, both sequentially and concurrently (in parallel), based on the Catania science gateway framework (CSGF).
3. Based on the CSGF, develop a framework for the efficient execution of ABS applications, both in sequence and parallel, based on the science gateway methodology.
4. Use the framework to develop an instance of a science gateway for ABS applications and test the feasibility of the framework. This will be done using an infection model that has been implemented using Recursive porous agent simulation toolkit (Repast) simphony.
5. Perform an evaluation on both the SG methodology and framework for executing ABS applications by applying to a much larger model such as an health economics model.



Solution Design

Science gateway layer: This layer consists of the different applications (represented as: infection model 1, model 2, and model 3) that has been ported and deployed on the science gateway.

Grid and cloud engine layer: This layer contains the grid and cloud engine with API/software. It has a science gateway interface to the portlets and a JSAGA API interface to the e-infrastructures.

DCI layer: This layer contains the e-infrastructures, such as cloud and grid, where jobs can be submitted for execution.

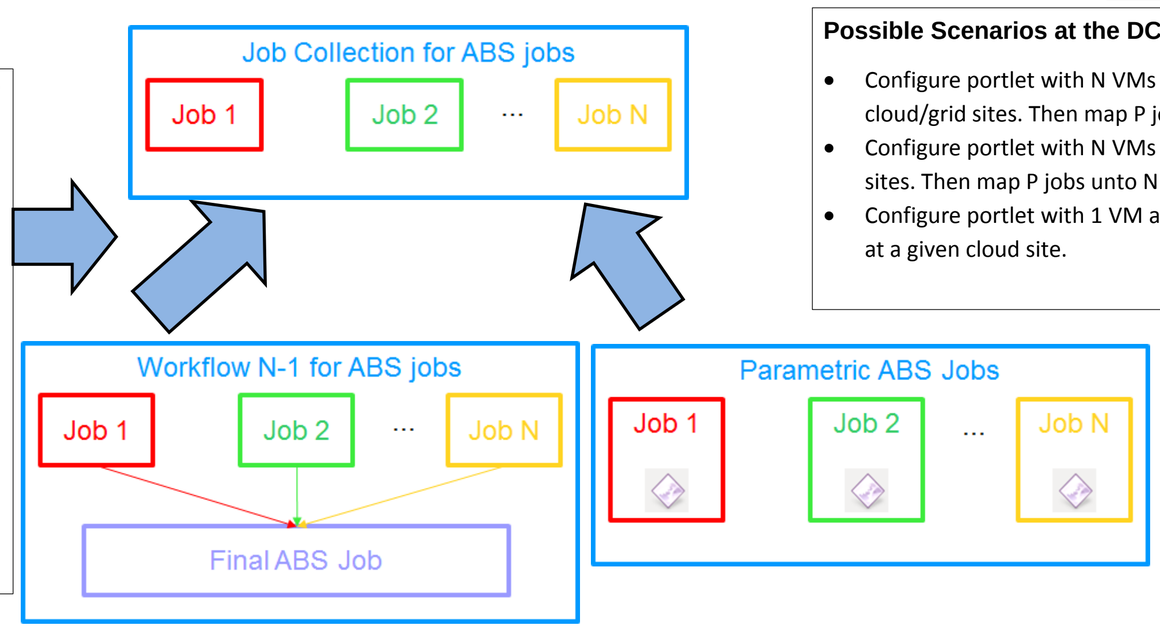
Possible Scenarios at the DCI level for ABS applications:

- Configure portlet with N VMs and 1 core each at different cloud/grid sites. Then map P jobs unto M VMs
- Configure portlet with N VMs and M cores each at different cloud sites. Then map P jobs unto N VMs and their M cores.
- Configure portlet with 1 VM and N cores (and OpenMP enabled) at a given cloud site.

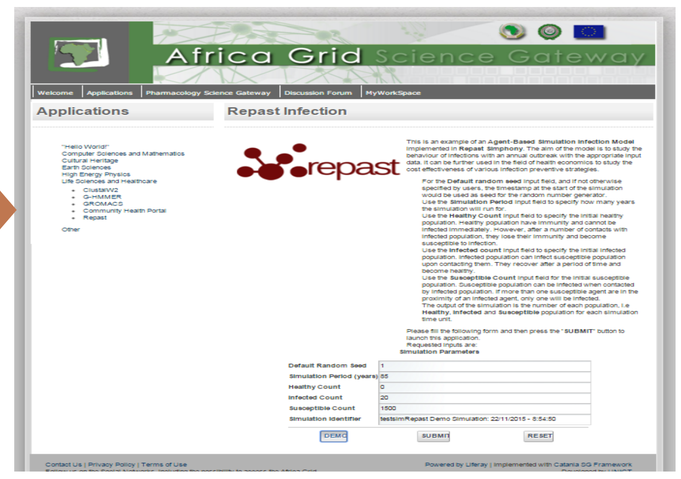
Solution Design

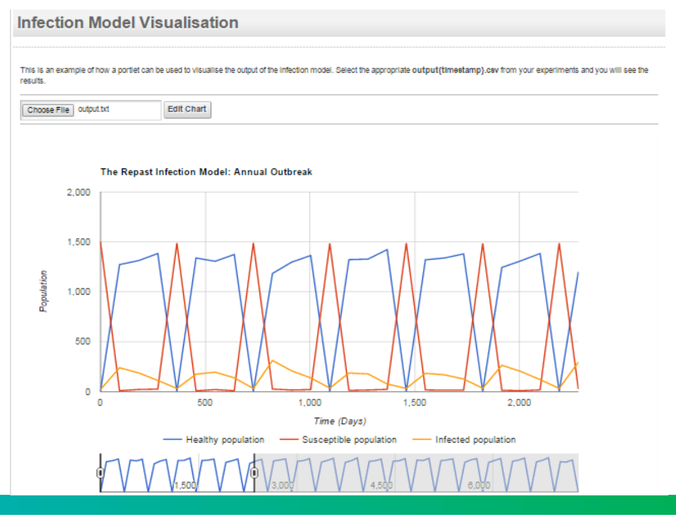
Possible Scenarios for parallelising our ABS application:

- **Jobcollection, Workflow and Job parametric.**
- **Make the Jobcollection class the superclass of both workflowN1 and JobParametric.**



Result





Contributions

The contribution of this research is fourfold:

- To develop a SG for the execution of ABS applications, on e-infrastructures, for life sciences and healthcare.
- To provide a systematic approach/framework of using a SG interface for the execution of ABS jobs, on e-infrastructures, that is specific to this domain of science.
- To use SGs and Open Access Data Repositories to support Open Science in order to enable scientists to easily access everything associated with the ABS application. (Simulation software, model, data, etc).
- Finally, it is anticipated that this framework will help in executing ABS applications in a reasonable amount of time.

References

Balasko, A., Farkas, Z. and Kacsuk, P. (2013) 'Building Science Gateways by Utilizing the Generic WS-PGRADE/GUSE Workflow System'.

Barbera, R., Fargetta, M. and Rotondo, R. (2011) 'A Simplified Access to Grid Resources by Science Gateways'.

Kocot, J., Szepieniec, T., Wojcik, P., Trzeciak, M., Golik, M., Grabarczyk, T., Siejkowski, H. and Sterzel, M. (2014) 'A Framework for Domain-Specific Science Gateways'.

Russell, M., Dziubecki, P., Grabowski, P., Krysinski, M., Kuczynski, T., Szjenfeld, D., Tarnawczyk, D., Wolniewicz, G. and Nabrzyski, J. (2008) 'The Vine Toolkit: A Java Framework for Developing Grid Applications'.