University of the Witwatersrand
Honours Course Research proposal

Continuous Delivery of Research Applications in a Distributed Environment

CODE-RADE

- Principal investigator (PI): Bruce Becker (CSIR Meraka Institute) - Senior Researcher, SANREN Competency Area; Coordinator, South African National Grid; Coordinator, Africa-Arabia Regional Operations Centre. (bbecker@csir.co.za)
- Host institution: Wits University

- Full title: Continuous Delivery of Research Applications in Distributed Environment
- Proposal short name: **CODE-RADE**
- **Proposal duration: 10 months**

1 Problem Statement

One of the aims of e-Infrastructures is to provide easy access to powerful computational and data platforms, to as many eligible users as possible. The South African National Grid (SAGrid), as part of the National Integrated Cyberinfrastructure System is no exception. While access to the users is being simplified greatly by the adoption of science gateways and identity federations, the community of application developers and technical support in scientific collaborations does not yet have an easy way to integrate these applications in the first place.

SAGrid has identified this as a gap in the services it provides and has developed a solution to the issue of easily integrating new applications into the infrastructure in a fast, flexible, distributed and reproducible way. Using existing tools and services, we have defined a simple set of tests which applications need to pass in order to be considered valid for the infrastructure. These can been encoded as automated tests using a continuous integration platform such as Jenkins. Interoperability between source code repositories, automated build systems, artifact creation and content delivery systems is crucial for the sustainability and uptake of the system.

The specific problems we aim to address in this limited-scope project are those of automation and distribution. Specifically

- How far is it feasible to automate the build steps in order to produce high-quality, redistributable artifacts ?
- How are dependencies of applications to be managed in various configuraitons ?
- How can Linux containers be used to distributed these artifacts to remote sites with the minimum intervention necessary by site administrators

2 Methodology

The project will work on a case-study basis, with applications of different architectures selected. Massively-parallel applications (such as GADGET), self-contained applications (user-provided code), applications with complex dependency trees (such as Quantum Espresso), and applications based on common frameworks (python or R applications).

These will be studied in terms of their requirements and execution models and the candidate will be expected to write the configuration of Jenkins jobs necessary to successfully build them. Appropriate methods of artifact creation should be studied and suggested, and then implemented. For example the candidate should investigate doing this with Linux containers.

Finally, the candidate should provide the (bash) module file necessary to execute the application on any site which subscribes to the CVMFS repositories.

3 Expected Results

It is expected that these case studies will lead to insights necessary to provide best-practice guildelines to future development and operation of the service, especially related to optimisation and automation procedures.

4 Resources and Support

The project has a solid base from which to start and the basic design and implementation has been presented at a few conferences already. The Jenkins instance has been operational for over a year already hosted by the University of the Free State at https://ci.sagrid.ac.za.

The project can count on the full support of SAGrid, and will have the resources of the Africa-Arabia Regional Operations Centre at its disposal. Where necessary, coordination with experts from peer infrastructures (in particular EGI.eu) will be brought to bear.

5 References

NICIS - The NICIS is a framework for the development of an integrated system for cyberinfrastrucutre in South Africa, including pillars of data, compute and network infrastructure. See http://www.dst.gov.za/index.php/resource-center/cyber-infrastructure2

Jenkins - Jenkins is a continuous integration platform widely used to automate testing and integration of applications. See http://jenkins-ci.org

EGI.eu - EGI.eu is a coordinating body for Europe's largest federated computing infrastructure. See http://www.egi.eu

Africa-Arabia Regional Operations Centre (AAROC) - AAROC is the regional body for African e-Infrastructure, coordinated by the CSIR Meraka Institute, and providing a platform for infrastructure interoperability between Africa and other regions, most notably Europe.