YABI: Connecting Australian HPC with the Cloud

Matthew Bellgard, Andrew Macgregor

Cenrte for Comparative Genomics, Murdoch University, Perth, Australia

mbellgard@ccg.murdoch.edu.au, amacgregor@ccg.murdoch.edu.au

GENERAL INFORMATION

Please answer the following questions to assist the conference organisers with scheduling.
- Is this workshop half-day or full-day? **Half a day**
- Who is the primary convener for the workshop? **Matthew Bellgard**
- Does the workshop include a hands-on component? **Yes**
- Are there any constraints on the number of attendees? **No**
- Are there any technical requirements beyond AV and access to wireless network? **No**

DESCRIPTION

There is an urgent need to have easy-to-use tools to enable Australian researchers to uptake national supercomputing data and compute infrastructure. These tools must cater for researchers that are non-HPC specialists and non-software progammers – the majority of potential eResearch users in Australia.

In this workshop, the open source system Yabi eResearch Tool is presented. Yabi primariliy enables seamless and transparent access to heterogenous HPC, Grid and cloud resources and then builds an analysis workflow environment that can create and reuse workflows. Yabi is able to manage large amounts of both raw and processed data in a secure and flexible way across geographically distributed compute resources. Users are able to either use the web-based environment to drag-and-drop tools to create sophisticated workflows or alternatively use the yabi-command-line, which is designed for users more comfortable writing scripts or for configuring external workflow environments. In existing workflow environments, configuring tools is a significant overhead. Yabi greatly simplifies this task by enabling system administrators to configure as well as manage running tools via a web-based environment and without the need to write/edit software programs/scripts.

Yabi is not tied to any single research discipline. Originally, Yabi was developed to leverage HPC investments in compute and storage for life scientsts in the areas of genomics, proteomics and metabolomcis. However, Yabi has proven to be applicable to other domains including radio astronomy and remote sensing. There are numerous local instalations of Yabi both in Australia and in other countries. Yabi may be found at: http://ccg.murdoch.edu.au/yabi/.

Reference: Hunter A, Macgregor B, Szabo T, Wellington C and Bellgard MI, Yabi: An online research environment for Grid, High Performance and Cloud computing, Source Code for Biology and Medicine 2012, 7:1 doi:10.1186/1751-0473-7-1 Published: 15 February 2012

OUTLINE

The Yabi workshop will cover the following topics.
1. **First topic.** *Transparent access to HPC and the Cloud via Yabi Introduction to Yabi and construction of basic workflows.* Topics covered include: Overview of Yabi; Using Yabi to access HPC and the Cloud; Designing and running workflows; Design Panel,;Tools Panel; Uploading files; Moving files between different resoures i.e. from HPC to Cloud, or local computer to Cloud; SSH, S3; Storing results, accessing files, avoiding lock-in; and Provenance.

   60 minutes

2. **Second topic.** *Administering Yabi.* Topcis include: adding execution and storage resources; managing users and credentials;adding tools, command line, R; worked examples, hostname, blast, rscript; importing and exporting tools; using webservices, automatically importing EBI webservices; monitoring workflows; debugging.

   60 minutes

3. **Third topic.** *Installing Yabi.* Topics include: downloading and installing Yabi on your own computer; testing with local execution and local filesystem; connecting your install to Cloud resources; installing and running for production using Apache; contributing and bug reporting.

60 minutes

WHO SHOULD ATTEND

- Researchers who wish to take advantage of disparate compute and storage resources from an intuitive web based application
- Researchers will gain an understanding of how it is possible to leverage HPC for their analysis without having to use a Linux command line
- System administrators of eResearch Tools will develop an understanding of how Yabi abstracts away the details of compute and storage resources for the user while still enabling the system administrator full control over the environment
- System administrators will gain an insight into what is required to deploy and maintain a Yabi installation

WHAT TO BRING

Attendees can bring a laptop.

ABOUT THE PRESENTERS

Matthew Bellgard. Matthew Bellgard is Murdoch University's Bioinformatics Chair and the Director of the Western Australian State Government Centre of Excellence, the Centre for Comparative Genomics. His scientific work has resulted in developments in the both the areas of pairwise sequence alignment and artificial intelligence, early detection of base composition differences in closely related bacterial species, whole genome sequence analysis and advances in the development of web-based integrated systems utilising high performance computing.

Andrew Macgregor. Andrew Macgregor is a software engineer focusing on Internet Application development using Python. Andrew has a degree in English and has also studied Computer Science. He also has an interest in cloud computing using the Amazon Web Services stack.

DELIVERY

Submit all extended abstracts as a Microsoft Word or PDF attachment to your submission via the author submission form for the conference: https://www.easychair.org/conferences/?conf=eresau2012. Your submission should also include a short abstract of no more than 300 words in the space provided in the submission form. Please see the call for participation at http://eresearch.edu.au/eres2012/participate for more information on the submission guidelines. Submissions are due by 8 June 2012.