Genomics Virtual Laboratory Workshop

Enis Afgan[1], Michael Pheasant[2], Andrew Lonie[3], Clare Sloggett[4]
[1]Victorian Life Sciences Computation Initiative, University of Melbourne, Australia, afgane@gmail.com
[2]University of Queensland, Brisbane, Australia, m.pheasant@uq.edu.au
[3]Victorian Life Sciences Computation Initiative, University of Melbourne, Australia, alonie@unimelb.edu.au
[4]Victorian Life Sciences Computation Initiative, University of Melbourne, Australia, sloc@unimelb.edu.au

General Information

Place the workshop title and presenter information above. Please answer the following questions to assist the conference organisers with scheduling.

- Is this workshop half-day or full-day? **Half-day**
- Who is the primary convener for the workshop? **Enis Afgan**
- Does the workshop include a hands-on component? **Yes**
- Are there any constraints on the number of attendees? **15-20 max**
- Are there any technical requirements beyond AV and access to wireless network? **Laptops required**

Description

The Genomics Virtual Laboratory (GVL) is a combination of scalable compute infrastructure, genomics analysis platforms, and support, all of which will be available at multiple physical nodes around Australia. Researchers can use one of the centrally managed instances of the GVL to directly perform genomic analyses and visualisation (with large resource/support allocations available through subscriptions); alternatively, research groups or institutes may instantiate, manage, and tailor their own instance of the genomics analysis platform on the same national cloud infrastructure, the NeCTAR Research Cloud (http://www.nectar.org.au/research-cloud).

The GVL is designed to scale to multiple locations and arbitrary cluster sizes on the Research Cloud through the CloudMan platform (http://usecloudman.org), and will be supported by comprehensive training courses, outreach programs and end-user support for subscribers.

At this stage of development, the GVL analysis platforms include: a prototype workflow management system based on the Galaxy framework (http://usegalaxy.org), a bioinformatics toolkit (http://cloudbiolinux.org), and a visualisation service based on the UCSC Genome Browser (http://genome.ucsc.edu), all implemented on the NeCTAR Research Cloud. In addition, a developing set of tutorials and exemplar workflows targeted at common high throughput genomics tasks will be available.

The aim of this workshop is to familiarise attendees with the instantiation, tailoring and use of a GVL instance on the Research Cloud.

Outline

1. **Overview of the GVL** in form of a presentation giving attendees an overview of the project, the scope, and the aims.
   15 minutes
2. **Instantiation of a small scalable GVL instance on the Research Cloud.** Attendees will learn how to instantiate a small private GVL instance on the NeCTAR Research Cloud using the CloudMan platform.
   30 minutes
3. **Tailoring of a GVL instance with new tools from the Galaxy Toolbox**. Attendees will learn how to add new tool functionality to the instantiated GVL.
   30 minutes
4. **Simple genomic analysis on the tailored GVL instance.** Attendees will run through a short workflow driven genomics practical exercise, exploring the analysis and visualisation characteristics of the GVL.
   90 minutes

Who Should Attend

This workshop is aimed at bioinformaticians and technical research support personnel who are interested in the Genomics Virtual Laboratory and scalable cloud data analysis platforms.

What to Bring

Attendees will need to bring a laptop with internet connectivity. All activities are web based so no specific software required beyond Firefox or equivalent. Attendees are expected to have some computing experience. Ideally, attendees would have some knowledge of or experience in bioinformatics, but it is not required. The nature of the genomics analysis task will be fully explained during the workshop.

About the Presenters

Dr Enis Afgan is a research scientist at the Life Sciences Computation Initiative (LSCC) of the University of Melbourne. He also holds a research scientist position at the Ruđer Bošković Institute (RBI) in Zagreb, Croatia. He obtained his Ph.D. in 2009 in grid computing from the Department of Computer and Information Sciences at the University of Alabama at Birmingham. His research interests focus around distributed computing, with the current emphasis on application- and user-level accessibility of cloud computing resources. Since 2009 he has been a member of the Galaxy Project team. He is the lead on the CloudMan and the Galaxy on the Cloud projects. These projects deliver complete (bioinformatics) analysis environments on cloud resources and enable domain scientists to perform needed computation without restrictions of locally available compute resources.

Dr Michael Pheasant is the manager for Genome Research Computing at UQ. He worked as a project scientist and software engineer on the UCSC Genome Browser staff in 2008 and 2009. He has published papers in genomics and received his PhD in Bioinformatics from the UQ in 2005 with Professor John Mattick's lab at the IMB. For many years previous to this he was the IT director of a software development company with offices in Australia and NZ.

A/Prof Andrew Lonie is a faculty member of the department of Computing and Information Systems at the University of Melbourne and foundation head of the Life Sciences Computation Centre, a cross-institutional centre of bioinformatics and computational biology expertise and infrastructure support within the Victorian Life Sciences Computation Initiative. Genomics, proteomics and biomedical image analysis experts within the LSCC collaborate with and support life sciences researchers in a variety of research projects across Victoria; the centre is also responsible for implementing and disseminating best practice methods and techniques, advising on experimental design and interpretation, and resourcing and maintaining informatics analysis platforms. A/Prof Lonie is also academic coordinator of the MSc (Bioinformatics) at the University of Melbourne.

Dr Clare Sloggett is a research scientist and bioinformatician at the Life Sciences Computation Initiative (LSCC) genomics node, at the Victorian Life Sciences Computation Initiative at the University of Melbourne. She obtained her PhD in computational physics from the University of New South Wales in 2007. She worked on research infrastructure at Intersect, the NSW eResearch body, before joining the LSCC in 2011. Her research interests are in the bioinformatics and biostatistics of high-throughput genomics and in understanding regulatory networks. She also has an interest in the architecture of workflow systems for flexible data analysis.