BIOINFORMATICS WORKSHOP ON
ENABLING DATA ANALYSIS WITH GALAXY CLOUDMAN

Grand Hotel Adriatic Convention Centre, Opatija, Adriatic Coast, Croatia
27th May 2014

Improve your bioinformatics productivity
Introduction to bioinformatics cloud service
Walk away knowing how to get to the infrastructure and the tools

Presenter

Enis Afgan
Ruđer Bošković Institute
Center for Informatics
and Computing
Zagreb, Croatia

Conference scope
Workshop is organized as a sister event of the DC VIS Conference held on MIPRO Convention 2014. The conference is devoted to presenting and exploring scientific and technological advancements and original innovative applications in the fields of Distributed computing, Visualization systems and Biomedical engineering (BME).

Conference web site
http://www.mipro.hr/MIPRO2014.DCVIS

General information

- Full-day workshop
- Workshop includes a hands-on component
- Maximum number of attendees is 40
- Technical requirements: Laptop

Description

Given the influx of digitized biological data, computational methods are becoming a required component of any experiment. However, computational data analysis methods impose obvious difficulties dealing with the requirement to provision, configure, and maintain computational and storage resources as well as the domain-specific tools. More subtly, computational methods impose difficulties with data sharing and experiment reproducibility. This workshop will demonstrate attendees how provision required computational resources on commercial and private cloud computing providers and offer promising solutions for dealing with data analysis, interpretation, and reproducibility.

Namely, the participants will establish a genomics platform on a computational cloud provider using CloudMan application. The established platform provides researchers with an accessible and scalable computational infrastructure that is preconfigured with genomic datasets and applications. Most notably, the platform includes a preconfigured Galaxy application, which allows researchers to access genomic datasets from multiple data sources and integrate them with own data to perform desired data analysis. A sample data analysis pipeline will be demonstrated along with how Galaxy enables reproducibility in bioinformatics.

Outline

1. **Overview of Galaxy and CloudMan applications** in form of a presentation giving attendees an overview of the domain and projects
   60 minutes
2. **Deploying a scalable Cluster-on-a-Cloud.** Attendees will learn how to deploy a small private virtualized clusters using the CloudMan platform.
   45 minutes
3. **Using Cluster-on-a-Cloud.** Attendees will learn how to run jobs on the Sun Grid Engine cloud cluster that underpins the GVL.
   45 minutes

4. **Overview of Galaxy application.** Attendees will learn notions of datasets, histories, workflows as well as how to upload private and public data.
60 minutes
5. **Bioinformatic runs and workflows.** Attendees will follow a sample bioinformatics analysis.
75 minutes
6. **Tailoring a Galaxy on the Cloud instance with new tools from the Galaxy Tool Shed.** Attendees will learn how to add new tool functionality to the instantiated Galaxy instance.
45 minutes
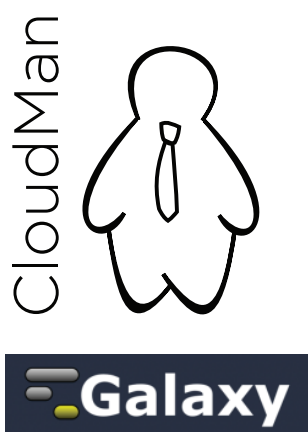
Who should attend

Researchers, bioinformaticians, technical research support personnel, and project managers interested in using novel eResearch cloud-based architectures for data analysis and experiment reproducibility.

What to bring

Attendees will need to bring a laptop with internet connectivity. All activities are web based so no specific software is required beyond Chrome 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 presenter

Dr Enis Afgan is a research scientist at the Ruđer Bošković Institute (RBI) in Zagreb, Croatia. He obtained his Ph.D. in 2009 from the University of Alabama at Birmingham (UAB). 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 and since 2012 a member of the Genomics Virtual Lab project. He is the lead on the CloudMan project; CloudMan delivers complete (bioinformatics) analysis environments on cloud resources and enables domain scientists to perform needed computation without restrictions of locally available compute resources.

This program may be subject to change.

Contact

For all future information please visit www.mipro.hr or contact directly Workshop organizers:

> Enis Afgan (Enis.Afgan@irb.hr)
> Karolj Skala (skala@irb.hr)

Conference operative:

> Jelena Čubrić
> Ruđer Bošković Institute
> Bijenička c. 54, HR-10000 Zagreb, Croatia
> E-mail: jelena.cubric@irb.hr
> GSM: +385 99 3833 888, fax: +385 1 4680 212