



# CRG CORE FACILITY TECHNOLOGY SYMPOSIUM

## ABSTRACTS

**9 November 2012**  
**09:30a.m. to 1:40p.m.**  
**PRBB Auditorium**

# CLOUD COMPUTING: DEMOCRATIZATION OF HIGH-THROUGHPUT DATA ANALYSIS

## Jesus ARTECHE

*IBM SmartCloud South Europe*

### Big Data on IBM SmartCloud

Genomics is now under a data explosion.

Hadoop has established itself as the preferred platform for genomic analysis.

Many tools and protocols have been developed in Hadoop that are readily available for researchers to deploy, with plenty of examples of how to use them. Many tools are also increasingly available in cloud-based services. Although cloud-based services might seem better suited for ad-hoc, short analysis, on small data sets, when coming to Hadoop they provide flexibility, scaling capabilities and quick access that may be the right choice for many Big Data situations and needs.

In this session Jesus, IBM SPGI Architect, describes Hadoop services in public cloud, their capabilities and costs.

## Ivona BRANDIĆ

*Vienna University of Technology, Vienna AT*

### Cloud Computing: Reducing Energy Consumption while Improving Bioinformatics?

With the rapid development of high-throughput technologies in recent years, huge amounts of data are being generated and stored in databases in the field of life science, which necessitates significant advances in computing capacity and performance. An example of such technologies is the high-throughput sequencing of the transcriptome (RNASeq), which has the potential to transform how gene structure and gene expression profiling are studied. Scientific workflow applications are crucial in enabling scientists to determine important information from those huge amounts of stored data.

On the other hand, Cloud computing represents a revolutionary approach for implementing energy efficient, cost efficient, and scalable ICT systems. Thereby, emerging virtualization technologies facilitate delivery of computational resources as services over networks. In this talk we discuss how the application of Cloud computing technologies can support the work of scientists working in the field of high-throughput sequencing, while at the same time optimizing utilization of resource and increasing energy efficiency.

## Konstantinos KRAMPIS

*J.Craig Venter Institute, Rockville (MD) US*

### Cloud BioLinux: open-source bioinformatics computing on the Cloud for genomics and beyond

Cloud BioLinux is an open-source Virtual Machine (VM), that enables scientists with access to high-performance computing infrastructures for bioinformatics using Clouds. Upon deployment of the VM, users have at their disposal a collection of pre-installed and configured software for alignment, clustering, assembly, visualization and editing of next-gen sequencing and other types of large-scale bioinformatic datasets. In order to provide a platform for software developers as well, we have included code libraries from the BioPython, BioPerl, R, and Ruby programming languages, in addition to a framework that creates on-demand, parallel computing clusters on the Cloud using multiple instances of the Cloud BioLinux VM.

End-users can simply access the bioinformatics tools using a graphical interface through a remote desktop connection. Furthermore, the extensive range of pre-installed software and code libraries provides a rich ecosystem for implementing scalable genomic data analysis pipelines, and use of the parallel computing capabilities of Cloud BioLinux on both commercial (Amazon) and private (Eucalyptus/OpenStack) Clouds.

This talk will provide an overview of the Cloud, and its potential as an informatics solution for researchers at academic and industrial laboratories in need of large-scale computing resources for genomic data analysis. We will also demonstrate how users can access the tools included with Cloud BioLinux, describe the fundamentals and options our platform offers for bioinformatics developers, while also provide examples of Cloud-based data analysis pipelines implemented at the J. Craig Venter Institute.

## Ryan SHUTTLEWORTH

*Amazon Web Services (AWS), Norwich UK*

### Amazon AWS and High Performance Computing: Concepts, Patterns and Practices

In this presentation Ryan Shuttleworth, Technical Evangelist for AWS, will introduce the AWS platform and how it applies to High Performance Computing. Discussing aspects of the platform that are relevant, and talking to customer examples, you'll hear how using AWS for complex workloads is a cost effective method for the application of 'disposable' supercomputing resources to complex computation problems.