

ResOps: Research DevOps across clouds

Thursday, 30 November 2017 11:30 (15 minutes)

As more and more researchers are hitting the limits of traditional computing facilities, the interest into using public and private clouds keeps growing. However, many researchers are unwilling and unable to leave behind their existing infrastructure meaning science is in a transitional period. At the moment, many research groups across Europe are struggling with the same basic questions when exploring the potential of cloud facilities: How do we move away from our traditional scheduling environment? How do we move and keep our data on the cloud? How do we run our workflows on the cloud efficiently and at low cost?

After a number of successful ventures at EMBL-EBI of both new and existing workflows into the cloud, we have used the lessons learned from these processes and combined them to create an interactive hands on workshop that aims to guide researchers getting their 'feet wet' for the first time with clouds. The basics of cloud computing are explained thoroughly, after which the case for moving towards cloud computing is made using benchmarks and real-life experiences. The course then dives deeper into the practical application of porting an existing HTC workflow to a cloud environment, discussing hybrid cloud setups where a traditional scheduler environment is emulated on cloud resources but also outlining how to move to a fully cloud-aware environment to make optimal use of resources available. Given this theoretical underpinning, participants are then able to get hands-on experience through a set of practicals which guide them through basic infrastructure deployment with the Terraform deployment tool and subsequent configuration of this infrastructure with tools such as Puppet and Ansible. They are then guided through emulating a traditional job scheduler resulting in a hybrid environment suitable for porting a traditionally scheduled workflow.

So far, four iterations of the course have been given and feedback has been overwhelmingly positive. More courses are being planned in which we hope to also reach non-bioinformatics researchers. This presentation will present the case for creating this course, give a summarized overview of the topics discussed and discuss further needs and potential additional training that is needed based on experiences and feedback from this course.

Topic Area

Data science and skills

Type of abstract

Presentation (15 minutes)

Primary author: VAN DEN BERGH, Erik (EMBL)

Co-authors: VIANELLO, Dario (EMBL); NEWHOUSE, Steven (EMBL)

Presenter: VAN DEN BERGH, Erik (EMBL)

Session Classification: Data science and skills presentations