



Cloud needs for training

Eija Korpelainen, WP₁₁



ELIXIR-EXCELERATE is funded by the European Commission within the Research Infrastructures programme of Horizon 2020, grant agreement number 676559.

www.elixir-europe.org/excelerate

Steven's wish for my talk today

"I would like to get a better understanding of what you would like to run on ELIXIR cloud resource:

- typical number of training sessions /week
- typical number of students in a class
- typical size & number of VMs created per student
- source of VMs for student [new VM per class, library of VMs, etc.]

If you have some experiences of using clouds as a platform for delivering training resources, that would also be interesting. Especially if you could identify where you had problems that you would like to see solved."



Why do we need cloud for training courses?

- Lot of software and reference data need to be installed
- Students need to have identical installation
- Students are bio/medical people
- Analysis jobs require CPU and memory
- 20-30 people run the analysis job at the same time
- The same course needs to be repeated in different cities / countries



Typical number of courses and students

- Roughly 1-2 bioinformatics training events / week in ELIXIR nodes altogether, but not all of these courses use cloud.
- Whether a course is run in the cloud depends on several things
 - Does the training topic covered require a lot of software and computing resources (e.g. NGS data analysis)?
 - Is there a suitable image (e.g. Rstudio) available?
 - Does the trainer know how to use cloud, is technical help available?
 - Is a cloud available?
- Typical number of students per course: 20-30

Typical size & number of VMs created per student

- It depends, there is nothing “typical”.
- R course: 4 cores and 16 GB memory is fine to run Rstudio for 24 students
- NGS course: 2 cores and 8 GB memory per analysis job. We use x number of VMs with 12 cores and 48 GB memory, depending on the number of students.
- Etc etc

Source of VMs [new VM per class, library of VMs, etc.]

- Typically the same image is used in several courses, but this depends on the course topic
 - For topics like NGS, the software and reference data needs to be updated often
- Marketplaces and registries of images are available (e.g. RAINbio)
- Trainers prefer images, admins prefer configuration management (e.g. Ansible)

Problems in using cloud for training

- Bioinformatics trainers lack the technical expertise needed
 - Cloud usage should be made easy (ELIXIR-FR has done a great job with IFB cloud)
 - Trainers need to be trained in using cloud
 - Technical help needs to be available
- It needs to be clear what clouds are available and how much does it cost
- Ideally students need access to identical setup after the course



Train-the-Trainer workshop “Using clouds and VMs in bioinformatics training” 23.-25.5 at CSC

- 33 bioinformatics trainers and technical specialists from 13 different countries
- Hands-on practicals with 4 clouds and Docker
 - IFB cloud, EGI FedCloud, cPouta cloud, AWS
- 5 presentations on other clouds including the Australian one
- Materials and presentation videos are available
 - <https://github.com/ekorpela/cloud-vm-workshop>
- Guideline document for trainers will be prepared



