Istanziazione e utilizzo di batch system on demand su infrastrutture Cloud. L'esempio pilota dell'esperimento AMS Welcome and Introduction

Daniele Spiga spiga@infn.it

🔗 eosc-hub.eu

🐦 @EOSC_eu

Dissemination level: Public/Confidential *If confidential, please define:*
Disclosing Party: (those disclosing confidential information)
Recipient Party: (to whom this information is disclosed, default: project consortium)

Welcome to Perugia

- This course is **jointly organized by INFN and EOSC-hub H2020 project**, as such it is part of the training program of the project

- We have 14 participants, mostly INFN staff, one external researcher (from AMS02 collaboration) and a remote participant (remote connection will be best effort)

  - People here have heterogeneous skills and background (*). The programme should, hopefully cover common requirements and needs

    - (*) we span from System Admins to end-user

- About the structure: **this course will have both theory and practice** (roughly half / half), more details later

EOSC-hub project

- This training course is supported also by the EOSC-hub project (https://eosc-hub.eu).

- EOSC-hub brings together multiple service providers to create the Hub: **a single contact point for European researchers and innovators** to discover, access, use and reuse a broad spectrum of resources for advanced data-driven research.

- The project mobilizes providers from the **EGI Federation, EUDAT CDI, INDIGO-DataCloud** and other major European research infrastructures to deliver a common catalogue of research data, services and software for research

About Rooms: we will use two rooms: IT-Lab (Flor-2) and Aula F (Flor-0)

- Due to some constraints and clash with teaching activities we can't use the IT-Lab all the time, however my plan is to optimize it's usage and this will imply we move a bit up and down:
- **Today** we will be here (**aula F**) till the coffee. Then we move on **IT-Lab**
- **Tomorrow and Wednesday** we will be always at **IT-Lab**
- **Thursday Morning** we will be in **Aula F** again

This should fit quite well with hands-on sessions.

- To be held in the lab

Logistic (cont)

- **Wireless access** is available via eduroam.
  - If you do not have eduroam, please contact us special we will provide credentials

- **Coffee breaks** are kindly provided by INFN Sezione di Perugia (Thanks!) and will be served at mid-morning and mid-afternoon
  - Today and Thursday outside of this room
  - Tuesday and Wednesday in the lab.

Note that **lunch is not provided**

• There are **several options** around here, bars, pizzerias and full restaurants (*).
  - In any case, please respect the timings (we planned 1h 30 for lunch break)

(*)
  - **100Dieci Caffè (self-service and bar)**
  - **Bar Restaurant Rubbagalline**
  - **Pizzeria La Casetta**
  - **Ecce Bombo Burger**
  - **Caffè Desìo**

Structure of the course

Refer to agenda at https://agenda.infn.it/category/23/ for details and timetable.
- All presentations will be uploaded there.

The objective of this training is to provide a comprehensive overview of the solutions recently implemented to allow the exploitation of Cloud based resources for "data analysis"
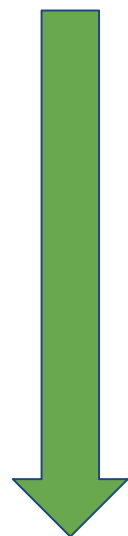
- We first describe main principles, the gensis and the technical architecture

- Then we'll walkthrough the building blocks, describing roles and responsibilities. We will practice also.

- This will bring us in showing the concrete use case of HTCondor batch system on demand

- We will then conclude with lectures on how to use the same strategy for diverse use cases

Today we start with a **general introduction** as well as an overview of the activities done in **the context of AMS02 collaboration**
- We will enter the technical part in the last hour: **The containers**

Tomorrow we will discuss the three pillars:
- **Container** (Docker) and orchestrations (K8s), **Automation** (Ansible), **Resource abstraction and Orchestration** (TOSCA)

Wednesday we will **connect the dots**:
- We will **generate our own cluster to run a HTCondor batch system**
- Focus on **usage of HTCondor**: from very basic to some advanced concepts

Thursday: **Going beyond HTCondor**, how to use the same strategy to address different types of workflows and use case

Additional Focuses

Managing Data:

- There will be a dedicated session on Data Management, cause this is a key aspect in almost all data intensive computation, and we will provide an overview of both technical and policies aspects

Managing Software:

- Software and user, or community, tailored code must be available in distributed computing. Here we rely on solution such as CVMFS and this will be treated, from user perspectives, as well.

The course is generally structured with lectures in and hands-on sessions, Hands-on will require access to testbeds and cloud resources:

- We will provide you access to a remote Virtual Machine which will be your user interface for all the exercise.
- The VM can be accessed from the IT-Lab terminal. If you prefer to use the laptop we can provide you with ssh-key (please ask)
    - Nothing prevent you to use your own laptop but this requires extra steps, also, in case of issues we can scratch the VM at any time.. While with laptop... :)

We will circulate detailed instructions
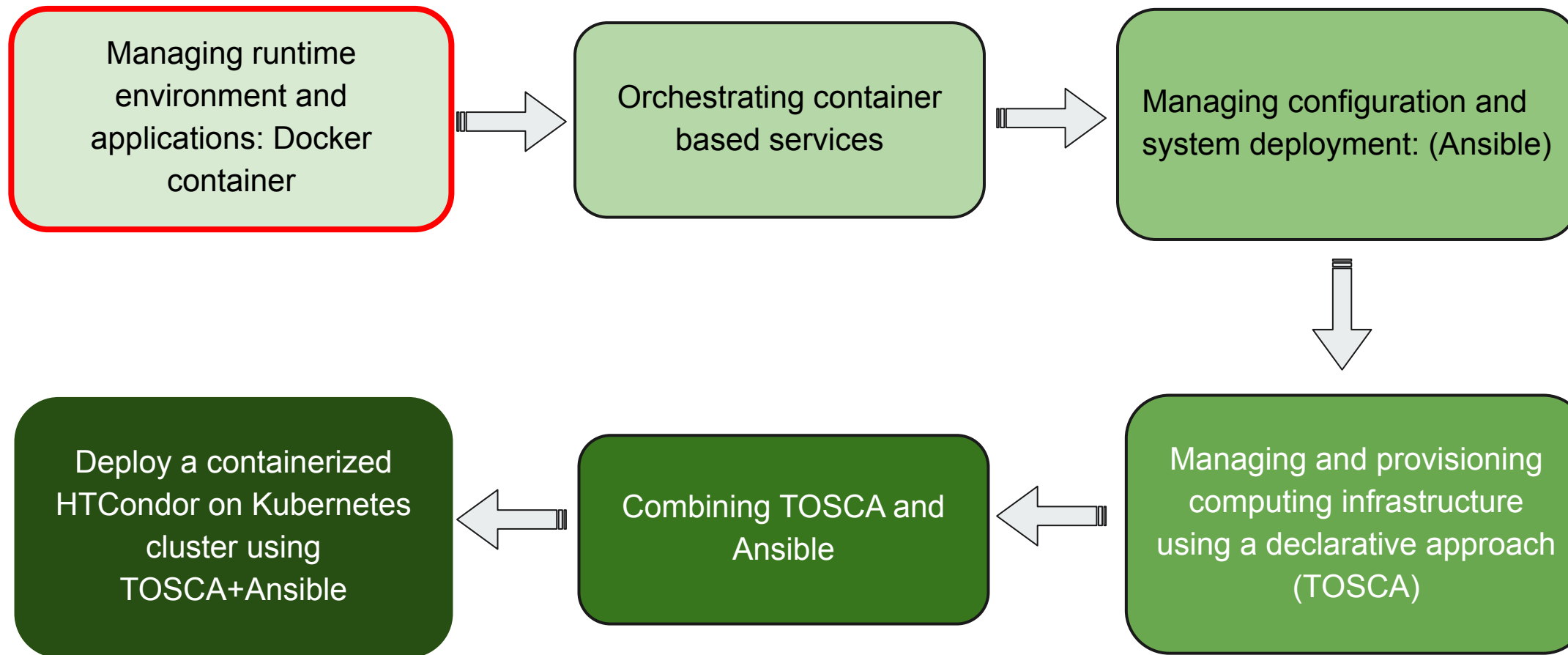
Speakers

This is the list of speakers:

- Marica Antonacci
- Daniele Cesini
- Diego Ciangottini
- Stefano Dal Pra
- Cristina Duma
- Davide Salomoni
- Daniele Spiga

Last but not least

A special thanks to all the speakers (+ Mirko) for their availability and their effort for the preparation of the material which include not only lectures but also

- Definition of exercises and setup of services and testbeds we will use during the school

Thanks also to Giacinto for all the contributions

Finally: a key to the success of the course is to keep it as much interactive as possible… please ask questions (in any language you prefer :))

- Agenda has been built to have some time "spare" on purpose, let's try to avoid stopping discussions etc.

Thank you for your attention!

Questions?

Contact

EOSC-hub

🔗 eosc-hub.eu   🐦 @EOSC_eu