RECAS-BARI: new high-level services for eScience researchers

Marica Antonacci$^1$, Giacinto Donvito$^1$, Gioacchino Vino$^1$, Michele Perniola$^2$

$^1$INFN Bari, $^2$Università degli Studi di Bari

EGI Conference 2020
2-5 November 2020 - Online
Outline

- ReCaS-Bari Overview
- Planned hardware upgrade
- Users and projects
- ReCaS Cloud Platform
  - High-level architecture
  - Objectives & challenges
- Use-cases
- Future work
ReCaS-Bari: the datacenter (the present)
Major hardware upgrade

EGI Conference, 2-5 November 2020 - Online
Cloud@ReCaS: Objectives

- Exploit the available resources efficiently
- Ensure isolation and security across different environments
- Provide users with easy-to-use tools
- Automate as much as possible (DevOps)
- Relieve the burden of System Administrators
ReCaS PaaS Platform

Based on the INDIGO/DEEP solutions to provide high-level interfaces to the end users:

- The **INDIGO Orchestrator** ensures the transparent access and exploitation of the resources, in the **cloud** and in the **HPC-like** environment.
- **RECAS-IAM** (based on INDIGO IAM) is used to manage the users identities and permissions.
Experimental phase: workflow involving “legacy” data access

Some users want to access the data they have already stored in the large parallel file-system based on GPFS and want to start an interactive session:

- The Platform is able to retrieve the user information (User ID, SSH public key) from the central store.
- The Platform deploys and configures the container and mounts the user home from GPFS.
- The user can access the container and/or its services (via ssh or http).
Already in production: workflow involving Cloud Object Storage

For some applications, the cloud storage is exploited to manage the user data in a transparent way:

- The platform uploads the data on behalf of the user on private Swift containers
- The job starts in a container configured to download the data from Swift and upload the generated output to Swift
- The user retrieves the generated output from the Object Storage
Job submission example

1. Configure and submit

2. Monitor and get output url

3. Download outputs data and view results

An equity-oriented rethink of global rankings with complex networks mapping development

Loredana Bellantuono 1, Alfonso Monaco 2, Sabina Tangaro 2, 3, Nicola Amoroso 4, 6, Vincenzo Aquaro 6, Roberto Bellotti 1, 2
Challenges

● Addressing the problems of shared resources
  ○ run a container with a user who has anything more than the minimum required privileges
● Enable additional security measure
  ○ Seccomp custom profiles
  ○ Tuning the capabilities available to containers
● Securing Container Secrets
  ○ secure store, such as HashiCorp Vault
Some use-cases

- Executing long-running services for data analytics with (or without) the usage of specialized hw (GPU, SSD, etc)
- Executing data processing job-like tasks
- Standard services, like: tensorflow, teano, Jupyter Notebook, etc
- Homemade sw using specialized libraries, that require building ad-hoc dockers
- Accessing large volume of data already stored in the large parallel file-system based on GPFS
Future work

• Provide a complete set of tools for the typical workflow:
  • create the custom docker, upload to a public or private repo, exploit it several time over the HPC cluster
  • Integrate the local cluster into the National federated Cloud infrastructure (INFN Cloud see talk “INFN-Cloud, an easy to use, distributed, user-centric Cloud infrastructure and solutions toolbox” -- 2 Nov 2020, 13:40)
Thank you
for your attention!