CAESAR service

A revolution in astronomy is unfolding. All-sky surveys performed by **next generation facilities**, such as the Vera Rubin Observatory or the Square Kilometer Array, will transform our understanding of the cosmos, allowing us to address the **most relevant astrophysical questions from a global perspective**. However, the unparalleled sky coverage, angular resolution and sensitivity of all these forthcoming surveys come with an unescapable drawback: the **increasing size and complexity** of the datasets will pose significant technical and scientific challenges, that will push the processing and analysis capabilities of the astrophysics community to the limits. To overcome such challenges, new approaches based on **distributed cloud architectures** and **powered by machine learning and deep learning** techniques will be required.

Much of the science to be done in this new, exciting era relies on a fundamental step, the **correct identification of astrophysical sources** in large scale maps. This task, traditionally done by means of **visual inspection** and cross-matching with existing catalogues, is now carried out by the so-called **source finders**, pieces of software specifically designed to identify and characterize astronomical sources in a **mostly automated way.** However, the new surveys are expected to reveal tens of millions of sources, reaching source densities of several 1000s per deg^2. These numbers will **render conventional source extraction techniques unfeasible**, exceeding the current capabilities of state-of-the-art source finders and underlining the need for a much **higher level of automation and knowledge extraction**.

But don't despair! **NEANIAS' response to this problem is the CAESAR** (Compact And Extended Source Automated Recognition) service, a powerful **tool particularly tailored to the needs of the SKA revolution**. CAESAR is able to extract and characterize sources from large astronomical images in an extremely efficient way. Unlike other state-of-the-art source finders, CAESAR is particularly capable of dealing **with the most extended, shallow sources**, like those that will likely be unveiled thanks to the unmatched capabilities of the SKA.

CAESAR service architecture

Initially developed as a command line tool, in the frame of the NEANIAS project CAESAR has evolved into a **cloud service**, dubbed CAESAR-REST. As depicted in Figure 1, CAESAR-REST relies on a **microservice-based architecture**, designed with scalability and performance in mind. The architecture comprises the following core components:

- A **REST API** based on Flask that constitutes the **main entry point of the service**. It provides endpoints to manage user data, submit and monitor jobs, and retrieve results.
- A **Job scheduler**, that handles the processing and execution of user submitted jobs, generally long-running tasks. Three different solutions are currently supported, namely *Kubernetes*, *Slurm* and *Celery*.

- **Storage components**. Operational data (quite heavy), like user images and job products, are stored in remote storage solutions like Network File System (NFS) and NextCloud. On the contrary, user-related information (lightweight), like file locations, descriptions and status of submitted jobs, are stored in a non-relational MongoDB database.

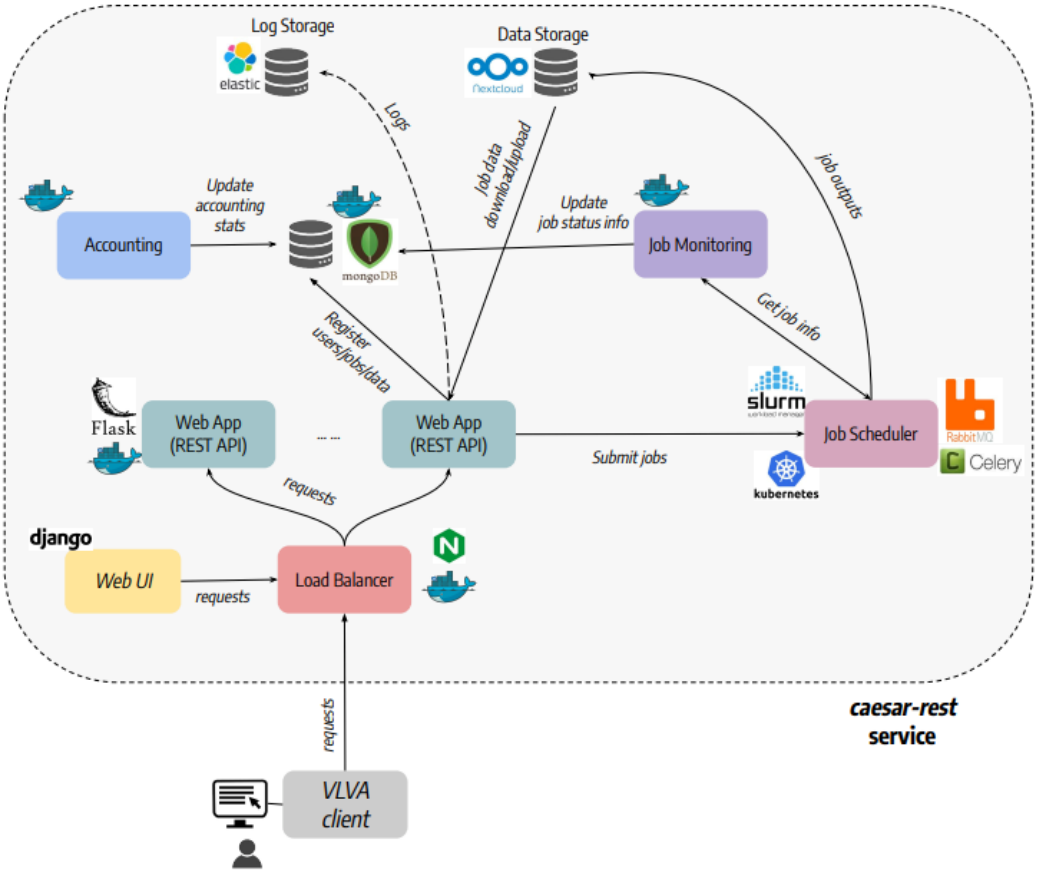

Figure 1. Schematic view of the service architecture of CAESAR-rest (Riggi et al. 2021)

These elements enable the typical workflow of CAESAR-REST user: *some astronomers, avid for knowledge, want to know which sources populate a certain patch of the sky; they upload the corresponding image, submit a source extraction job, and retrieve the results: a fully segmented map, along with a source catalogue containing all the desired information: source positions, angular sizes, fluxes and uncertainties.*

CAESAR-REST is a fundamental part of the NEANIAS Space service portfolio, and therefore it is **built on top of auxiliary NEANIAS core services**, in particular:

- The **AAI** service, a KeyCloak-based service for authentication and authorization of user requests.
- The **logging** and **accounting** services, ELK-based solutions for aggregating and processing application logs.

Finally, with the aim of **facilitating the usage of the service** in a user-friendly, intuitive way, CAESAR-REST is served through a **Django-based web interface**, providing seamless access to the main features of the service.

How to access the service

Currently, the CAESAR-REST architecture is deployed on a Kubernetes cluster operated by GARR. After being **successfully onboarded into EOSC**, the service is now accessible for the broad astrophysical community. The simplest way to access the CAESAR service is **through the web user interface** at https://caesar.neanias.eu. The user must log in using a valid Google or Microsoft account. From there, the user can navigate the left menu to access the different options:

Figure 2. Example of the CAESAR web user interface, showing a segmented radio image and the corresponding source catalogue (Riggi et al. 2021).

- The **manage data** view allows for managing the user files, i.e. uploading and downloading FITS files and assigning tags for easier identification.
- The **submit job** view allows for submitting source extraction jobs. The user can select the input image files on which the source extraction must be performed, tuning multiple options to their specific needs: how to deal with the image background, how to deal with extended sources, etc.
- The **check job** view allows for monitoring the status of the submitted jobs (that, depending on the image size, can take some time to complete). When a job finishes successfully, the user can access a previsualization of the segmented image and a simplified source catalogue, containing positional and flux-related information of the extracted sources, as displayed in Figure 2.

Input data format

At the moment, **CAESAR only supports FITS** (Flexible Image Transport System) files, the standard image format in most astrophysical fields. Even though CAESAR has been built to deal with the challenges of next-generation radio astronomical surveys, it is indeed a **domain-agnostic tool**, able to deal with infrared and optical images as well.

Current situation and future perspectives

In continuous development, CAESAR is fully operative and available through the EOSC marketplace. It is **currently being successfully employed to analyze the data** coming from **SKA precursors** such as ASKAP and MeerKAT.

CAESAR is envisaged as one of the cornerstones of CIRASA, a traversal platform that aims to provide the broad astronomical community with **efficient source extraction, data analysis and visualization capabilities**, encompassing state-of-the-art source finders, visual analytic tools and machine learning pipelines under a common architecture that can be replicated at a larger scale in the **future SKA Regional Centers**.

For **further details about CAESAR and the CIRASA platform**, don't miss the article "Astronomical source finding services for the CIRASA visual analytic platform" (Riggi et al. 2021), recently published in Astronomy and Computing.