Scalable and Distributed Processing of 3D Astronomical Data Cubes for Galaxy Evolution Studies Using the AC3 Framework

Mario Chamorro-Cazorla
Dpto. de Física de la Tierra y Astrofísica
Universidad Complutense de Madrid
Madrid, Spain
Email: mariocha@ucm.es

Cristina Catalán-Torrecilla
Dpto. de Física de la Tierra y Astrofísica
Universidad Complutense de Madrid
Madrid, Spain
Email: ccatalan@ucm.es

Ray Carroll
Waterford Research Interest Group
Red Hat Ireland
Waterford, Ireland
Email: rcarroll@redhat.com

Ben Capper
Waterford Research Interest Group
Red Hat Ireland
Waterford, Ireland
Email: bcapper@redhat.com

Ryan Jenkins
Waterford Research Interest Group
Red Hat Ireland
Waterford, Ireland
Email: ryjenkin@redhat.com

***Abstract*—We demonstrate the scalable processing of large-scale astronomical data using AC3's CECCM platform, showcasing real-time orchestration, intelligent microservice deployment, and efficient lifecycle management across the cloud-edge continuum for advanced astronomical data pipelines.**

I. Overview

This demonstration presents a working prototype of the third use case 'Deciphering the universe: processing hundreds of TBs of astronomy data' (UC3) from the AC3 (Agile and Cognitive Cloud-edge Continuum management) project. UC3 targets the scalable and automated processing of extremely large astronomical datasets, such as 3D spectral data cubes acquired from instruments like MEGARA [1], MaNGA [2] or MUSE [3]. These data contain rich spectral and spatial information essential for understanding stellar kinematics and populations. The demo showcases how AC3's Cloud-Edge Continuum Computing Manager (CECCM) enables distributed and automated deployment of microservices for data ingestion, transformation, and analysis in a Kubernetes-based infrastructure. Using the AC3 Ontology-based Semantic Reasoner (OSR) and Lifecycle Management (Maestro) components, applications are dynamically composed, deployed, and migrated across edge and cloud domains with zero-touch capabilities. Observers will see real-time resource allocation, workload balancing, and monitoring of key performance indicators via Prometheus.

The UC3 application architecture consists of two primary components: an Orchestrator that handles data ingestion and workload distribution, and a Processor that executes astronomical analysis tasks using containerized versions of domain-standard tools such as pPXF [4], Starlight [5] and STECKMAP [6]. Inter-component communication is implemented via RabbitMQ using AMQP messaging, ensuring decoupled, scalable, and reliable coordination across the cloud-edge infrastructure.

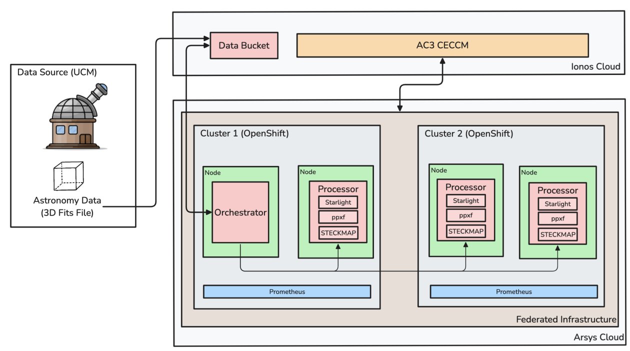

Fig. 1. Architecture diagram illustrating the integration of the UC3 astronomical data processing pipeline with AC3 components.

In this live demonstration, we start from a prepared astronomical data cube in FITS format, with sizes ranging from approximately 62 MB (MEGARA) to 32 GB (MUSE), from one of the mentioned instruments. Observers will see real-time resource allocation, data management and workload triggering via a custom GUI, workload balancing, and monitoring of key performance indicators through Prometheus.

The first processing step applies Voronoi binning to enhance the signal-to-noise ratio of the spectra and reduce the number of spectra to be analysed per observation. Voronoi binning is a technique that divides the spatial domain of the data cube into regions (Voronoi cells or polygons) based on proximity to a set of seed points. Each pixel is assigned to the region whose central point is closest to it, effectively grouping neighboring pixels with similar signal characteristics. This allows for adaptive spatial binning that preserves spatial resolution where

978-1-6654-6577-9/25/$31.00 ©2025 IEEE

the signal is strong and increases it where the signal is weak, optimizing the quality of the extracted spectra.

The second step involves decomposing the binned data cube into individual spectra files, each corresponding to a Voronoi region. These individual spectra are then analysed using the pPXF software to extract the kinematic properties of the stellar component of the observed galaxy, such as velocity and velocity dispersion, as well as other stellar parameters. Once the individual analyses are completed, spatial coherence is restored by reconstructing two-dimensional property maps, one for each derived parameter. These maps can be visualized as figures saved in PDF format and can also be collectively stored in a single FITS file for further analysis and archival. From here, we generate all necessary files for further scientific analysis using additional tools integrated in the UC3 pipeline, namely Starlight and STECKMAP. These tools perform full spectral fitting to derive the star formation history, chemical composition, and age distribution of the stellar populations, fixing the kinematics beforehand using the results from pPXF to avoid degeneracies in the fitting process. At the conclusion of the demo, attendees will observe how a reduced astronomical data cube is transformed, via fully automated orchestration, into a complete set of publishable scientific results, all managed through the AC3 CECCM framework across the edge-cloud continuum.

The use case exemplifies how CECCM can enable the astronomical community to scale up the analysis of vast and heterogeneous datasets. With support for hundreds of terabytes of data and dynamic, AI-driven service orchestration, UC3 demonstrates a breakthrough approach to managing data from current and future instruments, including observatories such as the James Webb Space Telescope (JWST). This capability empowers scientific and research teams to accelerate discovery and maximize the value of astronomical data through fully integrated, cloud-native computing.

II. Innovation

Traditional astronomical data processing systems rely on static, centralized infrastructures, struggling to manage the growing volume and complexity of datasets. UC3 introduces an AI-powered, modular framework for real-time, distributed processing over the cloud-edge continuum. Innovations include automated semantic-based application deployment, adaptive resource management using reinforcement learning, and seamless microservice migration across heterogeneous domains. The demo highlights the novel integration of federated CECC resources with explainable AI mechanisms to ensure SLA-compliant execution of astronomy workflows without human intervention.

UC3 also showcases the integration of legacy scientific software into a modern, containerized ecosystem, enabling reproducible and scalable analysis pipelines without modifying the underlying scientific codebases. By abstracting orchestration and infrastructure concerns through the CECCM platform, UC3 significantly reduces processing times for complex astronomical workflows, a critical advancement given the increasing data volumes from modern observatories. This efficiency gain not only accelerates scientific discovery but also demonstrates the practical impact and scalability of the AC3 framework in real-world, data-intensive research scenarios. Combined with AI-guided workload placement and full system resilience, UC3 sets a benchmark for next-generation astronomical data processing across hybrid cloud-edge environments.

III. Relevance

The IEEE SDN/NFV community will benefit from insights into dynamic service orchestration, intelligent workload migration, and zero-touch application management applied to large-scale scientific data. This demo exemplifies how SDN/NFV principles can be extended beyond telecommunications environments to high-performance, data-intensive research applications.

UC3 highlights the potential of cloud-native, AI-enhanced service orchestration for the astronomy domain, an example of how future SDN/NFV-based platforms can empower scientific discovery. The demonstration also showcases a complete, automated pipeline that could inspire similar workflows in domains such as medical imaging, environmental modeling, and physics-based simulations. By bringing together open standards, modular design, and cross-domain applicability, AC3's CECCM positions itself as a pioneering architecture for orchestration in the age of data-driven science.

It also illustrates how CECCM supports reproducible research and FAIR (Findable, Accessible, Interoperable, Reusable) data principles by enabling streamlined, automated workflows that directly support astronomers in producing scientifically publishable results.

Acknowledgment

The authors acknowledge the support of AC3, a project funded by the European Union's Horizon Europe Research and Innovation programme under grant agreement No 101093129.

References

[1] Gil de Paz, A. et al., "MEGARA, the new intermediate-resolution optical IFU and MOS for GTC: getting ready for the telescope", in Ground-based and Airborne Instrumentation for Astronomy VI, 2016, vol. 9908, Art. no. 99081K. doi:10.1117/12.2231988.

[2] Drory, N. et al., "The MaNGA Integral Field Unit Fiber Feed System for the Sloan 2.5 m Telescope", The Astronomical Journal, vol. 149, no. 2, Art. no. 77, IOP, 2015. doi:10.1088/0004-6256/149/2/77.

[3] Bacon, R. et al., "The MUSE second-generation VLT instrument", in Ground-based and Airborne Instrumentation for Astronomy III, 2010, vol. 7735, Art. no. 773508. doi:10.1117/12.856027.

[4] Cappellari, M. and Emsellem, E., "Parametric Recovery of Line-of-Sight Velocity Distributions from Absorption-Line Spectra of Galaxies via Penalized Likelihood", Publications of the Astronomical Society of the Pacific, vol. 116, no. 816, IOP, pp. 138–147, 2004. doi:10.1086/381875.

[5] Cid Fernandes, R., Mateus, A., Sodré, L., Stasińska, G., and Gomes, J. M., "Semi-empirical analysis of Sloan Digital Sky Survey galaxies - I. Spectral synthesis method", Monthly Notices of the Royal Astronomical Society, vol. 358, no. 2, OUP, pp. 363–378, 2005. doi:10.1111/j.1365-2966.2005.08752.x.

[6] Ocvirk, P., Pichon, C., Lançon, A., and Thiébaut, E., "STECKMAP: STEllar Content and Kinematics from high resolution galactic spectra via Maximum A Posteriori", Monthly Notices of the Royal Astronomical Society, vol. 365, no. 1, OUP, pp. 74–84, 2006. doi:10.1111/j.1365-2966.2005.09323.x.