A System for Distributed Computation Task Management in Containers

Bc. Kristián Kováč, xkovac61@stud.fit.vutbr.cz

Supervisor: RNDr. Marek Rychlý, Ph.D. Consultant: RNDr. Lukáš Hejtmánek, Ph.D.

BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

MUNI ICS

European Genomic Data Infrastructure

Excel@FIT 2025

Motivation

In the field of rare and cancer diseases, collaboration across Europe is playing an increasingly important role. The ability to identify patients with the same disease and similar genomic profiles is a key concept behind the GDI project, which aims to create an environment for collaborative search and analysis of genomic data across much of Europe. To achieve this, the project must overcome numerous challenges in building a secure, federated infrastructure that enables controlled, scalable, and user-friendly genomic data analysis.

Genenomic Data Infrastructure services

Users discover datasets and request access via the User Portal service. Authorization and identity management are handled securely by REMS and Life Science AAI, ensuring controlled data access to Sensitive Data Archives (SDA). Datasets remain safely stored in SDA; users explore data availability using the Beacon service, while authorized compute backends securely retrieve data through the standardized HTSGET protocol.

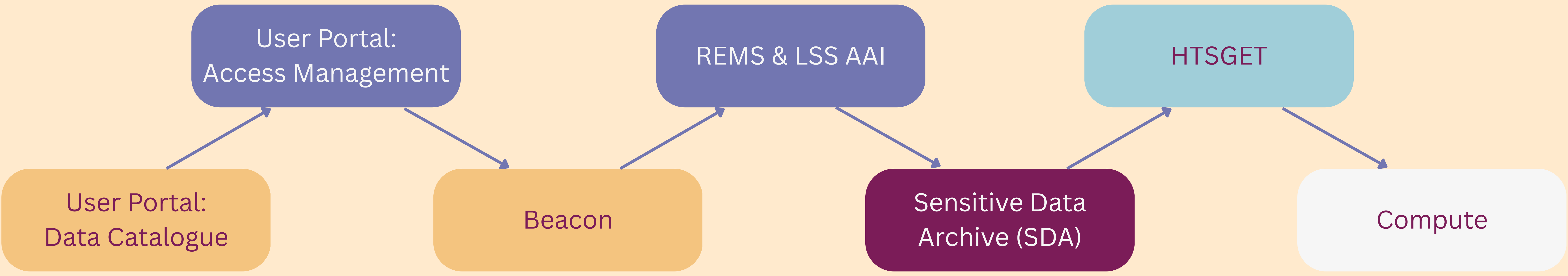
User Portal: Access Management
REMS & LSS AAI
HTSGET
User Portal: Data Catalogue
Beacon
Sensitive Data Archive (SDA)
Compute

Figure 1: Genomic Data Infrastructure services

Federated architecture

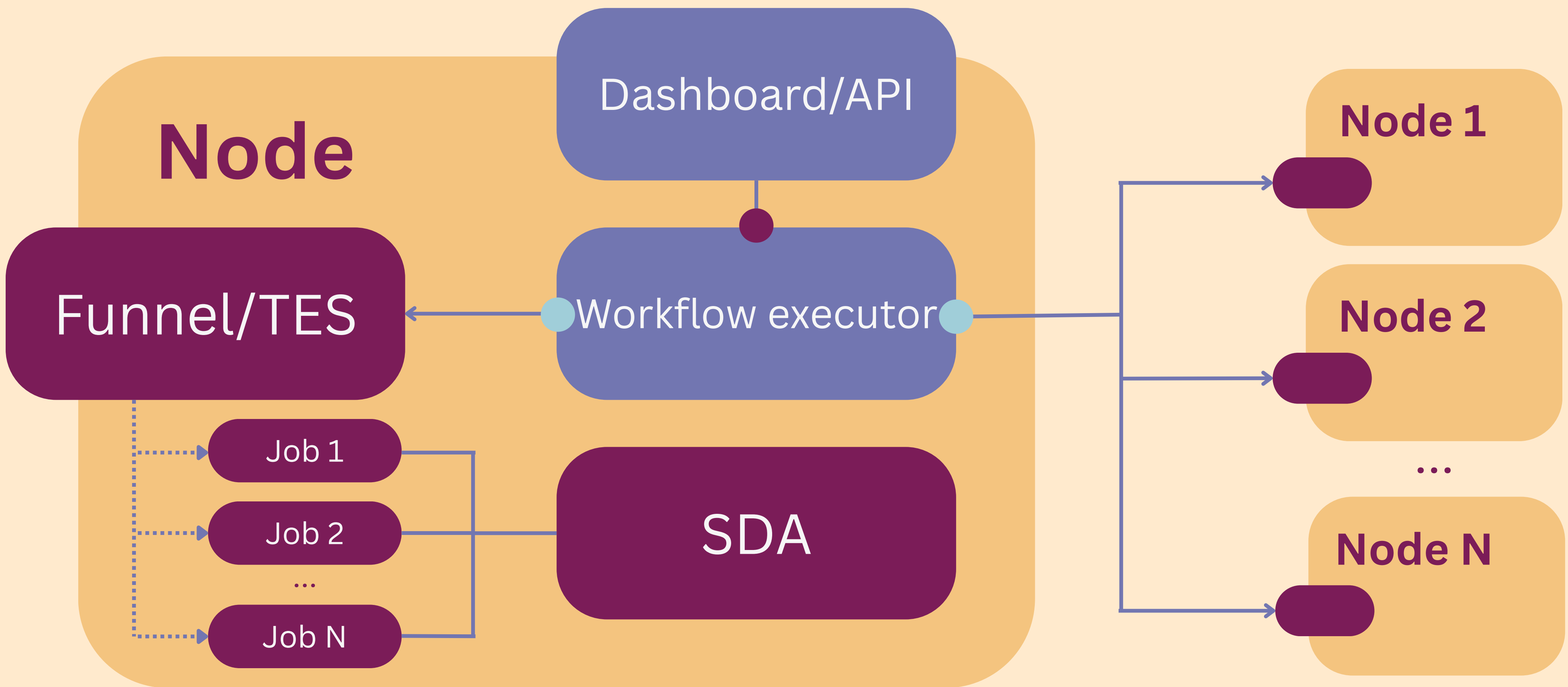

Figure 2: The Genomic Data Infrastructure (GDI) employs a federated architecture composed of multiple nodes, each independently handling secure genomic data computation, enabling collective yet decentralized execution of sensitive genomic workflows across institutional or national boundaries.

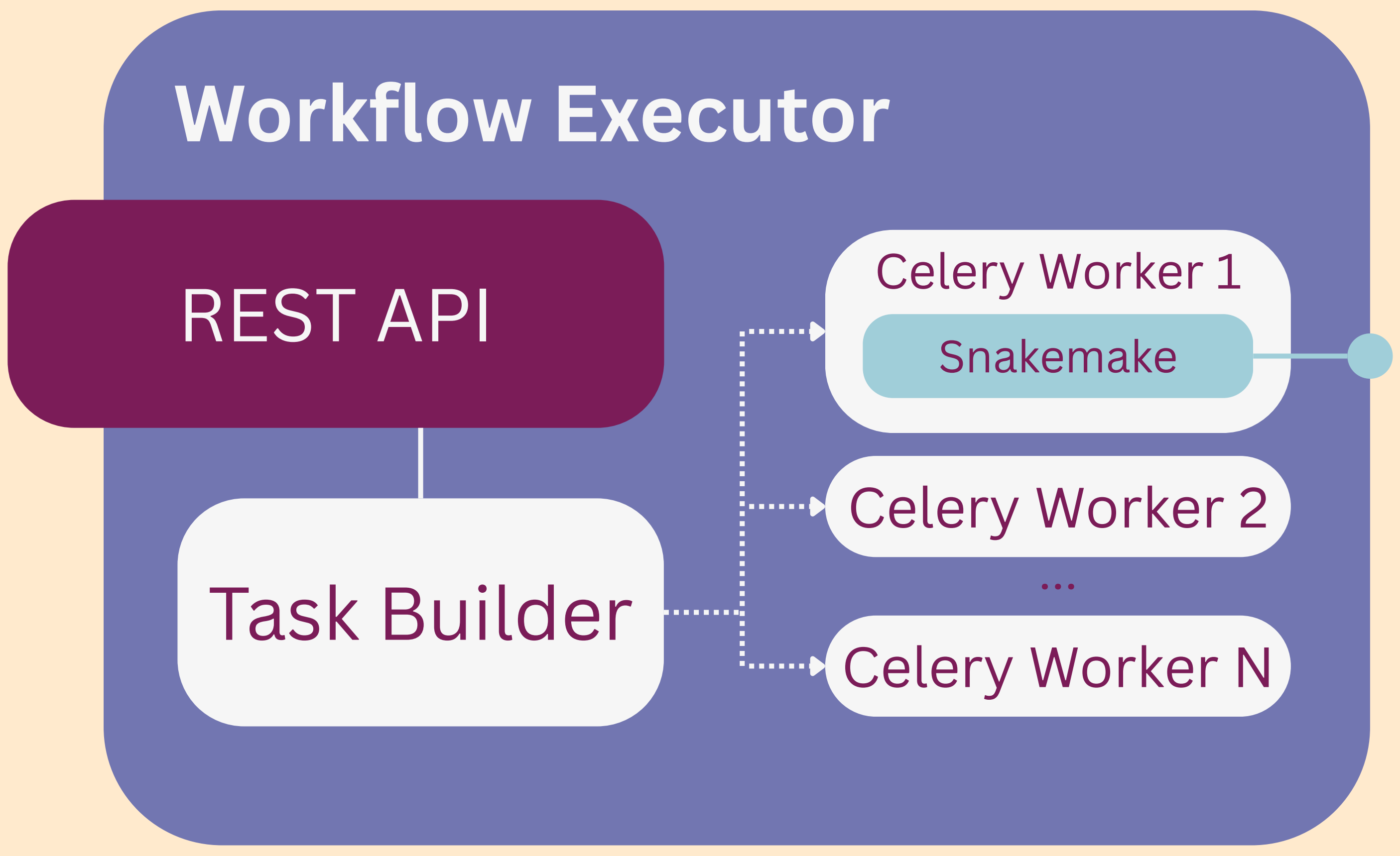

Figure 3: The workflow executor provides stable management of computational tasks by combining the scalability of Python's Celery framework with the workflow orchestration capabilities of Snakemake.