![SovereignEDGE.eu COGNIT logo]

A Cognitive Serverless Framework for the Cloud-Edge Continuum

D4.7 COGNIT Serverless Platform - Software Source - b

Version 1.0

30 April 2024

Abstract

COGNIT is an open-source serverless framework for the cognitive cloud-edge continuum that enables the seamless, transparent, and trustworthy integration of data processing resources from providers and on-premises data centers in the cloud-edge continuum. Automatic and intelligent adaptation to where and how data is being processed according to application requirements can be achieved by enabling an AI-Enabled orchestrator. Further, the orchestrator is able to optimise the placement, acts when changing in application demands and behaviour, and the operation of the infrastructure in terms of the main environmental sustainability metrics. This document provides a catalogue of those open source software components developed in WP4 "AI-Enabled Distributed Serverless Platform and Workload Orchestration" during the Second Research & Innovation Cycle (M10-M15) as part of the implementation of several of the main components of the COGNIT Framework (i.e. Cloud-Edge Manager and AI-Enabled Orchestrator).

Deliverable Metadata

Project Title:	A Cognitive Serverless Framework for the Cloud-Edge Continuum
Project Acronym:	SovereignEdge.Cognit
Call:	HORIZON-CL4-2022-DATA-01-02
Grant Agreement:	101092711
WP number and Title:	WP4. AI-Enabled Distributed Serverless Platform and Workload Orchestration
Nature:	R: Report
Dissemination Level:	PU: Public
Version:	1.0
Contractual Date of Delivery:	31/03/2024
Actual Date of Delivery:	30/04/2024
Lead Author:	Monowar Bhuyan (UMU) & Paul Townend (UMU)
Authors:	Simon Bonér (UMU), Malik Bouhou (CETIC), Zhou Zhou (UMU), Aritz Brosa (Ikerlan), Idoia de la Iglesia (Ikerlan), Sébastien Dupont (CETIC), Aitor Garciandia (Ikerlan), Joan Iglesias (ACISA), Tomasz Korniluk (Phoenix), Johan Kristiansson (RISE), Antonio Lalaguna (ACISA), Marco Mancini (OpenNebula), Alberto P. Martí (OpenNebula), Philippe Massonet (CETIC), Nikolaos Matskanis (CETIC), Daniel Olsson (RISE), Per-Olov Östberg (UMU), Goiuri Peralta (Ikerlan), Samuel Pérez (Ikerlan), Bruno Rodríguez (OpenNebula), Juan José Ruiz (ACISA), Kaja Swat (Phoenix), Thomas Ohlson Timoudas (RISE), Iván Valdés (Ikerlan), Constantino Vázquez (OpenNebula), David Carracedo (OpenNebula), Ignacio M. Llorente (OpenNebula), Victor Palma (OpenNebula), Michal Opala (OpenNebula), Pavel Czerny (OpenNebula), Jackub Walczak (OpenNebula).
Status:	Submitted

Document History

Version	Issue Date	Status[1]	Content and changes
0.1	20/04/2024	Draft	Initial Draft
0.2	26/04/2024	Peer-Reviewed	Reviewed Draft
1.0	30/04/2024	Submitted	Final Version

Peer Review History

Version	Peer Review Date	Reviewed By
0.1	24/04/2024	Antonio Álvarez (OpenNebula)
0.2	26/04/2024	Goiuri Peralta and Aritz Brosa (Ikerlan)

Summary of Changes from Previous Versions

First Version of Deliverable D4.7

---

[1] A deliverable can be in one of these stages: Draft, Peer-Reviewed, Submitted, and Approved.

Executive Summary

This is the second "COGNIT Serverless Platform - Software Source" report that has been produced in WP4 "AI-Enabled Distributed Serverless Platform and Workload Orchestration". It provides a short description, licence, version, code repository and user guide, as well as the design, testing, and verification references of each of the software requirements that have had active development tasks during the Second Research & Innovation Cycle (M10-M15) in connection with these main components of the COGNIT Framework:

Cloud-Edge Manager

- **SR4.3** Serverless Runtime Deployment:

  Deploy Serverless Runtime as Virtualized Workloads (e.g. Containers or VMs/microVMs) on the cloud-edge infrastructure.

- **SR4.4** Metrics, Monitoring, Auditing:

  Edge-Clusters monitoring, Serverless Runtimes metrics collection and continuous security assessment.

AI-Enabled Orchestrator

- **SR5.2** Smart Deployment of Serverless Runtimes:

  Implement a Smart Workload Orchestrator (SWO) that exposes a REST API used by the Cloud-Edge Manager for requesting the deployment plans used for provisioning the Serverless Runtimes.

- **SR5.3** Scheduling Mechanisms:

  Implement an interference-aware scheduler that places the Serverless Runtimes on the Edge-Clusters resources according to the deployment plan provided by the AI-Enabled Orchestrator.

This deliverable has been released at the end of the Second Research & Innovation Cycle (M15), and will be updated with incremental releases at the end of each research and innovation cycle in M21, M27, and M33.

Table of Contents

Abbreviations and Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
CPU	Central Processing Unit
FaaS	Function as a Service
REST	Representational State Transfer
SR	Serverless Runtimes
VM	Virtual Machine

1. Cloud-Edge Manager

[SR4.3] Serverless Runtime Deployment	
Description	The Serverless Runtime is the main management unit of the COGNIT Framework. It is defined by a document that is sent by the Device Client to the Provisioning Engine that communicates with the Cloud-Edge Manager (in particular the OpenNebula OneFlow[2] component) which allows it to handle both FaaS and DaaS as a single entity. In this second cycle, improvements in the Contextualization mechanism of OpenNebula were made to ensure that the Serverless Runtime is active when reported to the Device Client.
Licence	Apache 2.0
Version	OpenNebula 6.8.2
Design	D4.2 → [SR4.3] Serverless Runtime Deployment
Code	Public Repository
User Guide	Repository README / Serverless Runtime update guide
Testing	D5.3 → 10.4 Cloud-Edge Manager
Verification	D5.3 → 10.4 Cloud-Edge Manager

[SR4.4] Metrics, Monitoring, Auditing	
Description	Metrics collected by monitoring systems provide valuable information on the operational efficiency, resource utilisation and sustainability of data centres and serverless environments. In this second cycle, the focus has been on COGNIT specific metrics (Serverless Runtime metrics needed integration with Prometheus in OpenNebula), energy metrics (in particular, energy consumption per VM, to complement energy consumption per host) and geolocation metrics (for the hypervisor hosts).
Licence	Apache 2.0
Version	1.0
Design	D4.2 → [SR4.4] Metrics, Monitoring, Auditing

---

[2] https://docs.opennebula.io/6.8/management_and_operations/multivm_service_management/index.html

Code	COGNIT Release 1.0
User Guide	Repository README
Testing	D5.3 → 10.4 Cloud-Edge Manager
Verification	D5.3 → 10.4 Cloud-Edge Manager

2. AI-Enabled Orchestrator

SR5.2 Smart Deployment of Serverless Runtimes	
Description	The AI-Enabled Orchestrator is responsible for making intelligent decisions in the COGNIT framework. It interacts with the Cloud-Edge manager to provide smart scheduling of serverless runtimes based on historical data, optimising for an energy-efficient and green-energy in serverless runtimes placement.
Licence	Apache 2.0
Design	D4.2 → [SR5.2] Smart Deployment of Serverless Runtimes
Code	Public Repository
User Guide	Repository README
Testing	D5.3 → 10.5 AI-Enabled Orchestrator
Verification	D5.3 → 10.5 AI-Enabled Orchestrator

SR5.3 Scheduling Mechanisms	
Description	An interference-aware scheduler has been implemented and works as an external scheduler, where it decides the suitable host according to the available green-energy host. However, this scheduling is integrated with the default OpenNebula scheduler, where there is filtering on the available nodes to remove those that are not suitable for a particular VM (this can happen for multiple reasons, for instance, not enough capacity in terms of CPU and/or memory).
Licence	Apache 2.0
Version	Interference-Aware Scheduler
Design	D4.2 → [SR5.3] Scheduling Mechanism for Smart Deployment of Serverless Runtimes
Code	Public Repository
User Guide	OpenNebula official documentation
Testing	D5.3 → 10.5 AI-Enabled Orchestrator
Verification	D5.3 → 10.5 AI-Enabled Orchestrator