

D4.8 COGNIT Serverless Platform - Software Source - c

Version 1.0

11 November 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 Third Research & Innovation Cycle (M16-M21) as part of the implementation of several of the main components of the COGNIT Framework (i.e. Cloud-Edge Manager and AI-Enabled Orchestrator).



Copyright © 2023 SovereignEdge.Cognit. All rights reserved.



This project is funded by the European Union’s Horizon Europe research and innovation programme under Grant Agreement 101092711 – SovereignEdge.Cognit



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

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:	30/09/2024
Actual Date of Delivery:	11/11/2024
Lead Author:	Monowar Bhuyan (UMU) & Paul Townend (UMU)
Authors:	Yashwant Singh Patel (UMU), Malik Bouhou (CETIC), Javad Forough (UMU), Aritz Brosa (Ikerlan), Idoia de la Iglesia (Ikerlan), Sébastien Dupont (CETIC), Fátima Fernández (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), , Mikel Irazola (Ikerlan), Álvaro Puente (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	30/10/2024	Draft	Initial Draft
0.2	08/11/2024	Peer-Reviewed	Reviewed Draft
1.0	11/11/2024	Submitted	Final Version

Peer Review History

Version	Peer Review Date	Reviewed By
0.1	06/11/2024	Antonio Álvarez (OpenNebula)
0.1	08/11/2024	Mikel Irazola (Ikerlan)

Summary of Changes from Previous Versions

First Version of Deliverable D4.8

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

Executive Summary

This is the third “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 Third Research & Innovation Cycle (M16-M21) in connection with these main components of the COGNIT Framework:

Cloud-Edge Manager

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

The Cloud-Edge Manager must be able to deploy Serverless Runtimes as Virtualized Workloads within an Edge Cluster.

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

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

- **SR4.5** Authentication & Authorization:

Authentication and authorization mechanisms for accessing cloud-edge infrastructure resources by the devices for offloading workloads.

AI-Enabled Orchestrator

- **SR5.1** Building Learning Models:

Provide AI/ML models based on collected metrics from the Cloud-Edge Manager monitoring service related to Edge Clusters, Serverless Runtimes, and infrastructure usage.

- **SR5.2** Smart management of Cloud-Edge resources:

AI-Enabled Orchestrator is responsible for managing and optimizing the lifecycle of Edge Clusters and serverless runtimes within Edge Clusters according to the application requirements, infrastructure and virtual resource usage, and energy-aware policies.

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

Table of Contents

Abbreviations and Acronyms	5
1. Cloud-Edge Manager	6
2. AI-Enabled Orchestrator	8

Abbreviations and Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
CPU	Central Processing Unit
FaaS	Function as a Service
FFNN	Feed-Forward Neural Network
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
REST	Representational State Transfer
SR	Serverless Runtimes
TCN	Temporal Convolutional Network
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. Its life cycle is controlled by the Cloud/Edge Manager. In this development cycle we have developed and deployed a particular network configuration of the Cloud/Edge Manager to allow for cross-site migrations of Serverless Runtimes. This network configuration is applied by OpsForge on top of a stock OpenNebula install.
Licence	Apache 2.0
Version	OpenNebula 6.8.2 OpsForge 2.0
Design	D4.3 → SR4.3 Serverless Runtime Deployment
Code	OpenNebula Public Repository OpsForge COGNIT Repository
User Guide	OpsForge README
Testing	D5.4 → 9.4 Cloud-Edge Manager
Verification	D5.4 → 9.4 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 third cycle, the focus has been placed on analysing the current Scaphandre integration to extend it so energy consumption processes within the Serverless Runtime are reported to the AI-Orchestrator.
Licence	Apache 2.0
Version	2.0
Design	D4.3 → [SR4.4] Metrics, Monitoring, Auditing
Code	COGNIT Release 2.0

User Guide	Repository README
Testing	D5.4 → 9.4 Cloud-Edge Manager
Verification	D5.4 → 9.4 Cloud-Edge Manager

[SR4.5] Authentication & Authorization	
Description	Authentication and authorization mechanisms for accessing cloud-edge infrastructure resources by the devices for offloading workloads. A new integration with Biscuit has been carried out to satisfy this software requirement and allow for JWT-based authentication
Licence	Apache 2.0
Version	2.0
Design	D4.3 → SR4.5 Authentication & Authorization
Code	COGNIT Release 2.0
User Guide	Repository README
Testing	D5.4 → 9.4 Cloud-Edge Manager
Verification	D5.4 → 9.4 Cloud-Edge Manager

2. AI-Enabled Orchestrator

[SR5.1] Building Learning Models

Description The AI-Enabled Orchestrator utilises machine learning models developed (e.g., LSTM, GRU, FFNN, TCN) to predict workload behaviour and resource usage. The prediction results are combined with scheduling algorithms to optimise the serverless runtimes across cloud-edge resources, enhancing energy efficiency and green-energy usage. The orchestrator leverages historical data to refine its learning models adaptively.

Licence Apache 2.0

Version 2.0

Design D4.3 → [SR5.1] Building Learning Models

Code [Public Repository](#)

User Guide [Repository README](#)

Testing D5.4 → 9.5 AI-Enabled Orchestrator

Verification D5.4 → 9.5 AI-Enabled Orchestrator

[SR5.2] Smart Management of Cloud-Edge Resources

Description The AI-enabled orchestrator is responsible for managing and optimising the lifecycle of Edge Clusters and serverless runtimes within Edge Clusters according to the application requirements, infrastructure and virtual resource usage, and energy-aware policies.

Licence Apache 2.0

Version 1.0

Design D4.3 → [SR5.2] Smart Management of Cloud-Edge Resources

Code [Public Repository](#)

User Guide [Repository README](#)

Testing D5.4 → 9.5 AI-Enabled Orchestrator

Verification D5.4 → 9.5 AI-Enabled Orchestrator