EUROPEAN ALLIANCE
FOR INDUSTRIAL DATA,
EDGE AND CLOUD

TELCO CLOUD REFERENCE ARCHITECTURE

MAIN CONTRIBUTORS

Alfonso Carrillo Aspiazu (OpenNebula Systems), Manuela Bargis (Telecom Italia), Juan Carlos Garcia (OpenNebula Systems), Andreas Florath (Deutsche Telekom), Madalin Neag (OpenNebula Systems), Guillaume Nevicato (Orange), Toon Norp (TNO), Rosaria Persico (Telecom Italia), Charles Schulz (Vates), Kai Steuernagel (Deutsche Telekom), Luis Velarde (Telefónica)

EDITORS

Juan Carlos García (OpenNebula Systems), Dimosthenis Kyriazis (University of Piraeus)

ABOUT THE ALLIANCE & THE WORKING GROUP

The **European Alliance for Industrial Data, Edge and Cloud** [1] brings together businesses, Member States' representatives, and relevant experts to jointly define strategic investment roadmaps to enable the next generation of highly secure, distributed, interoperable, and resource-efficient computing technologies. The work is facilitated by the European Commission's Directorate-General for Communications Networks, Content and Technology (DG CNECT).

The **Cloud - Edge Working Group** (WG) brings together the main European Industry players in cloud computing that compiled the European Industrial Technology Roadmap for the Next-generation Cloud-Edge in 2023 [2]. Following the release of the aforementioned roadmap, as well as the publication of a thematic roadmap on cloud environments for the telecommunications industry (referred to as "Telco Cloud") [3], specific WG members have focused on the delivery of a reference architecture for Telco Cloud, presented in this document. The members of the WG that have co-authored this reference architecture are the following:

Deutsche Telekom
OpenNebula Systems
Orange
Telefonica
Telecom Italia
TNO
Vates

The WG is chaired by Ignacio M. Llorente (OpenNebula Systems), co-chaired by Bjorn Hakansson (TNO) and Maria Barros Weiss (IONOS), and facilitated by Ana Juan Ferrer (European Commission). The group that developed the current thematic roadmap is chaired by Juan Carlos García (OpenNebula Systems).

EXECUTIVE SUMMARY

The multi-provider cloud-edge continuum that Europe is targeting needs to be based on an architecture and technology that meets the requirements of complex applications, such as cloud-native telecommunication services and other critical services.

This document introduces a Reference Architecture for the aforementioned cloud-edge continuum from a functional perspective, describing building blocks and information flows across these blocks.

This Reference Architecture builds on top of mature virtualisation and containerisation technologies, that compose the virtualisation layer, adding a multi-provider management layer that allows to deploy applications over a distributed multi-provider, multi-technology cloud-edge infrastructure, thus eliminating the dependency and lock-in risks that current public cloud services present.

As new components, the Reference Architecture defines:

- The Multi-Cloud Orchestrator, that creates, upgrades and removes virtual resources (virtual machines or container clusters) over heterogeneous environments, while also facilitating the deployment of applications on the virtual resources, streamlining cloud and application lifecycle management and optimizing the use of these resources.
- The Multi-Cluster Managers and Virtual Infrastructure Platform Managers that abstract the underlying virtualisation or

containerisation technology providing a uniform interface, towards enhanced interoperability and portability.

- The Telco Cloud Connectivity Manager that provides seamless, end-to-end connectivity across edge and cloud environments, effectively covering all segments: public cloud networking, private cloud networking and the public network connecting them.

Some initiatives are already contributing to the development of different components of the architecture. These contributions as well as current gaps are also presented in this document.

It should be noted that the current architecture needs to be complemented with the development of other aspects to realise its implementation, i.e. integration, exposure and federation solutions.

TABLE OF CONTENTS

INTRODUCTION

A Telco Cloud is a scalable and elastic pool of shareable physical or virtual resources that meet the requirements of network functions and services. It follows the cloud computing paradigm, i.e. access through a network, with self-service provisioning and administration on demand [4]. A Telco Cloud incorporates the distributed computing systems (compute and storage), the software-defined communications networks that link systems together and the mechanisms to deploy and manage workloads (implementing network functions and other telco capabilities) across the distributed edge to cloud environment.

The Telco Cloud further to hosting network functionality (network functions associated with mobile core, RAN, SDN controllers, fixed access, etc.), may host other types of complex applications and services. It should be noted that an application is considered to be complex when it comprises multiple software components, microservices or modules (workloads) that are connected to each other and may reside in a heterogeneous set of distributed computing nodes, the edge-cloud continuum. Telecom services are examples of these complex applications, while others refer to AI pipelines (both for internal use like network optimisation and to be provided "as a Service" to customers), edge applications and intelligent software agents (representing persons, devices, drones, etc.). Mobile networks are evolving to a sixth generation that considers all these capabilities as part of the 3GPP system.

This document introduces a Telco Cloud Reference Architecture (TCRA), which describes a system that provides the distributed edge-cloud environment, built on a set of heterogeneous infrastructure, required to host complex applications or services, and the mechanisms to deploy them in a seamless way in that environment.

The Telco Cloud has specific requirements emerging from the characteristics of the network technology (5G, 6G, fibre broadband, satellite, etc.) it hosts, the hosting third-party applications (AR/VR, remote machine control, AI inference…) that may be collocated and the distributed nature of the networks from edge to central clouds, which demand certain features beyond the ones offered by typical cloud computing systems.

These requirements were described in the *Telco Cloud thematic roadmap* [3] and were previously introduced at the *European Industrial Technology Roadmap for the Next-Generation Cloud-Edge* [2]. As a summary:

- Specific infrastructure configurations, to ensure proper performance and efficiency of some demanding network functions. Requirements such as: (a) enhanced performance, including EPA capabilities such as Multus-DPDK, SR-IOV, CPU-pinning or Hugepages, (b) support for PTP synchronisation, (c) hardware acceleration (SmartNICs, GPUs, FPGAs…), (d) bare metal support, (e) virtual machine support covering both virtualisation and containerisation (e.g. kubevirt.io),

and (f) specific form factor, environmentally robust, lightweight servers to fit in space- and energy-constrained sites.

- Operation support technologies to manage the highly distributed computing infrastructure and the chains of network functions deployed on top. The infrastructure life cycle management requirements include: (a) bare metal provisioning automation, (b) multi-cluster monitoring, (c) networking automation, (d) remote automation and control and limited on-site intervention, (e) infrastructure customisation (e.g. hardware acceleration) with blueprints covering the most frequent configurations, and (f) multi-cloud orchestration, managing container clusters in different public and private clouds, over several cloud technologies.
- Automation and orchestration approaches to manage the life cycle of the vast and highly interlinked set of workloads (network functions and/or application components) that are required to deliver a network service or any other distributed complex application. The set of requirements in this context includes: (a) dynamic application orchestration, (b) edge discovery and mobility management to adapt the Telco Cloud and network/edge support to the user mobility, (c) federation to enable edge-cloud compute sharing and neutral compute host scenarios, (d) portability between different cloud technologies and providers based on standard functionality and APIs, (e) multi-tenancy and isolation, (f) security and regulatory compliance, and (g) dynamic resource allocation.

Although not specific to Telco, there are other aspects to be considered in the TCRA:

- Technology-agnostic integration across multiple cloud and edge environments, both private and public, enabling an efficient and sustainable operation.
- Management of aspects like contracts, accounting and pricing in a hybrid multi-cloud Telco Cloud, towards the optimisation of workload placement and the control of cost efficiency.
- Availability, metering and monitoring, and a proper SLA management.
- Portability and interoperability of data, systems and services in order to facilitate Telco Cloud adoption by telecom operators, since they are essential to provide the necessary confidence in moving data and services across *multiple cloud* environments and reduce lock-in risks.
- Security and Privacy towards a high level of confidence, as the computing and storage infrastructure are critical pillars for the *network* that need to rely on a trustworthy and reliable system.
- Fulfilment of the requirements of future applications that currently have a strong tendency towards even smaller micro-services, increased usage of PaaS, serverless technology and portability.

The TCRA is aligned with the EU digital strategy [5] focusing on:

- *Digital sovereignty,* avoiding dependency on foreign entities and regulations for digital technologies and services at different levels, including hardware, software, and communications.

- *Strategic autonomy* supported by a diverse supply chain. Virtualisation and container orchestration solutions need to be hardware-agnostic and provide interoperability and portability to avoid lock-in.
- *Sustainability,* considering environmental aspects when managing resources like hardware or energy.
- *Trustworthy data processing* features to guarantee interoperability, data protection, portability, transparency, reliability and trusted data sharing among companies.
- *Cybersecurity,* higher levels of security and an increased operational capacity to prevent and manage attack scenarios are fundamental in critical infrastructures supporting essential services.

The proposed TCRA must be able to support the scenario of an edge-cloud continuum composed of thousands of edge nodes served by multiple providers, set as a target in EU's Digital Compass for 2030.

Overall, the objective of Telco Cloud Reference Architecture is the following:

- Definition of the Telco Cloud, the associated computing services and functional components, the guiding principles for the design and evolution of its reference architecture, and the essential features of a Telco Cloud system.
- Description of the characteristics of Telco Cloud systems, defining:
    - The basic functionalities and their corresponding components, and the relationships to each other and to the environment (external systems, other Telco Clouds and customers).
    - The organisation of components in layers and/or domains.
- Provision of a vendor-neutral reference definition of the Telco Cloud, with the following aims:
    - Consistency with the requirements of the Telco Industry (mainly, but not exclusively, those defined in the Telco Cloud thematic roadmap) and with the EU principles (e.g. sovereignty, sustainability, security, etc.).
    - Creation of a level playing field for the industry to discuss and compare different Telco Cloud initiatives and offerings.

The TCRA does not include detailed technical specifications, does not prescribe a specific technical implementation, and does not limit innovation.

This document also identifies projects, open-source communities, research programmes and other initiatives that develop or have developed TCRA components and solutions, and approaches that could contribute to their implementation.

SECTION 1: TELCO CLOUD REFERENCE ARCHITECTURE OVERVIEW

The Telco Cloud system is in charge of deploying, upgrading or removing the service according to specific requirements, expressed in terms of list of service components to be deployed, upgraded or removed, as well as of relationship among those components. This is accomplished using the design features of the system, which allow the structuring of a service function chain in such a way as to deliver the service. The aforementioned service components may be VAFs (i.e. Virtual Application Functions, as application components that are configured to be deployed on virtual machines, with an example being the Virtual Network Functions - VNFs in a telco service), CAFs (i.e. Containerised Application Functions, as application components that are configured to be deployed in k8s containers, with an example being the Containerised Network Functions - CNFs in a telco service, also named cloud-native network functions by CNCF), or PAFs (i.e. Physical Application Functions, as application components that are deployed in a specific dedicated hardware or appliance, with an example being the Physical Network Functions - PNFs in a telco service).

The Telco Cloud system should support a sequence of actions required to deploy a service:

- For each service component described in a service function chain (i.e. a graph that represents the connections between the different service components that are required for them to deliver the service), the system:
    - Selects the deployment location (based on the component requirements, hardware availability and area to be covered), checks the availability of the required VMs (VAFs) and/or container clusters (CAFs) and the necessary hardware capacity at the selected location and, in case needed: (i) allocates physical resources, bare metal servers and other hardware - if not already available, (ii) assigns these physical resources to virtual machines - if they are required (request to deploy VAFs or CAFs over VMs) and they do not exist or should be upgraded, and (iii) assigns the virtual machines or bare metal servers to container clusters when required (deployment of CAFs) - if the cluster is not created or needs to be scaled up.
    - Deploys the service component (create an instance at the container cluster/VM).
- The system then chains the service components to deliver the telco service and configures the connectivity to deliver the expected performance.

In this context, the architecture of the Telco Cloud system presented in this document is described by: (i) *decomposition in functional components or building blocks*, ensuring that each component is described in terms of what it does and what it requires (e.g. resources, infrastructure configuration,

software packages, etc), and (ii) *workflows* that capture actions and interactions between different functional components, showing how the system and its components should behave when there is the need to deploy, upgrade or remove an application or service.

Architecture principles

The new services require the architecture to be dynamic in time and location, guaranteeing that workloads (the software enabling the services) are deployed at the right place with the right size at all times. To achieve the aforementioned aspects, the proposed architecture should consider the following principles:

- Modular, based on components with well-defined functionality connected by well-defined interfaces/methods for interactions.
- Open, interfaces are standardised and publicly available.
- Multi-site, distributed along a heterogeneous set of edge and cloud computing nodes.
- Full support for cloud-native and future applications.
- Sustainable, consuming resources and energy only when and where needed.
- Secure-by-design.
- Best practices on cloud application design.
- Optimised communication to certain components such as databases or persistent storage, balancing flexibility and performance.
- Ability to provide access to specific hardware (CPUs, neuronal accelerators, etc) for acceleration, security, confidentiality and performance.

1.1 Architecture overview

The following capabilities are offered by the proposed architecture in order to cope with a dynamic and multi-site environment:

- Workload placement decision: This capability is able to identify, out of all the existing sites in the distributed infrastructure topology, which ones are the right sites to deploy a given workload at a given time. This choice may be based on the user location, acquired for instance by the Multi-Cloud Orchestrator from the 5G Core's NEF function, and other parameters such as the application requirements (e.g. latency, bandwidth, etc.) or the context (e.g. load of the network, load of the compute nodes, etc.).
- Infrastructure configuration and management: Capability enabling the on-demand setup of the infrastructure required to host the workloads.
- Workload deployment: Once the infrastructure is ready, this capability performs the actual on-demand deployment of the software when it is required and removes it when it is no longer needed.

- Network configuration: The use of the services relies on the data network provided by the Operator. It does not only deploy software in the right location but also allows the data to flow in and out of that site and get to the user and to other application components. This capability performs the required modifications in the network to allow the correct data transfer.
- Workload portability: This capability enables the software to be moved between sites if required, in order to guarantee the objective of "having the workloads *deployed at the right place/with the right technology at any time*".

1.2 Main building blocks

The *building blocks* of the reference architecture control different aspects of service and infrastructure deployment, management and orchestration over the Telco Cloud.

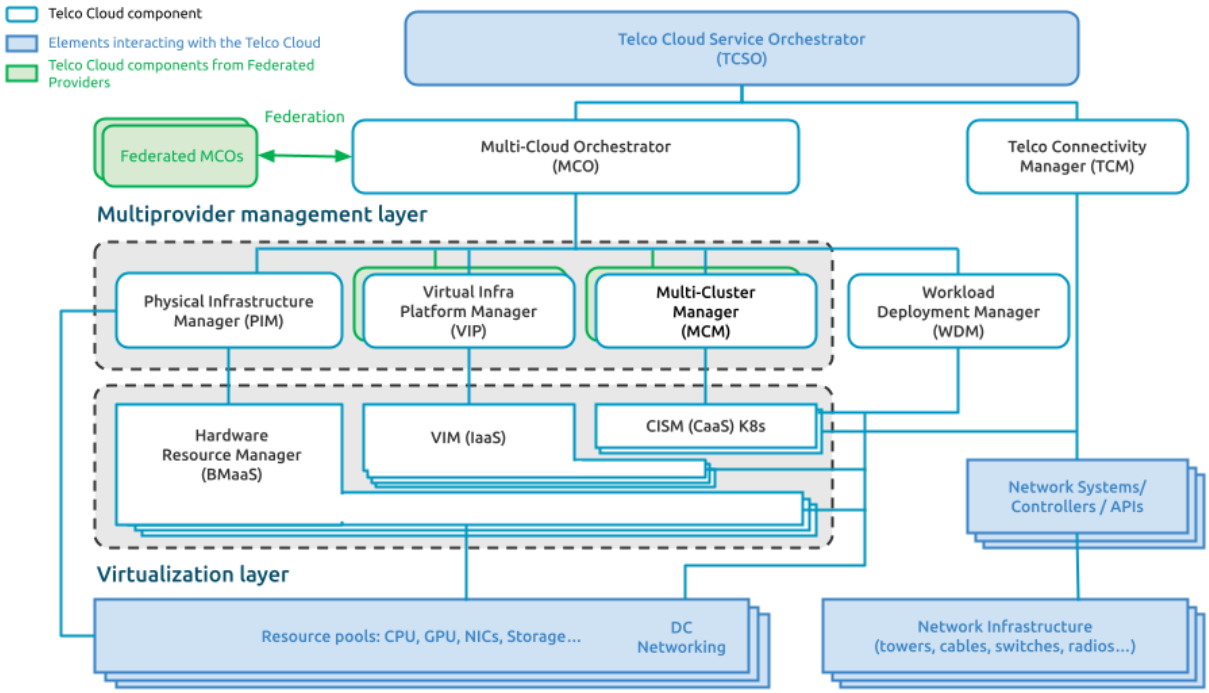

Figure 1: Telco Cloud Reference Architecture diagram

Telco Cloud Service Orchestrator

The Telco Cloud Service Orchestrator (TCSO) represents the highest level of abstraction that covers the deployment and lifecycle management of a complete *Telco Cloud-based service*. It allows users to define, deploy and manage a telecom or application service based on a chain of cloud-based network functions or application components (namely "workloads"). It is comprised of the following:

- A *design function* that allows users to define the service based on components (workloads) and workflows.
- *Software lifecycle management function* that manages the full lifecycle of application components via CI/CD and GitOps procedures. This is a critical support cross-function to allow for effective collaboration of vendors, infrastructure providers and telco cloud customers (e.g. telecom operators).
- A *deployment and management function* that is automated.

The deployment and management function relies on the *Multi-Cloud Orchestrator* (MCO) to generate the necessary cloud resources and deploy the workloads, and on a *Telco Connectivity Manager* (TCM) that provides the necessary connectivity between the workloads and the user device or application.

Even if the TCSO is considered out of the scope of the Telco Cloud system, it represents the entity that triggers the actions over the Telco Cloud. A Telco Cloud can be orchestrated and managed without this component, but the actions over MCO and TCM will need to be triggered manually, which is not uncommon today. In the case of complex services, the lifecycle management and orchestration require a high number of human interventions, making the process slow, inefficient and prone to errors, and a TCSO becomes highly recommended.

Multi-Cloud Orchestrator

As part of its functionality, the Multi-Cloud Orchestrator (MCO) receives requests to deploy a workload with a set of attributes (performance, area of service, containerised/virtualised, etc.). Based on those attributes:

- The MCO identifies the location(s) to deploy the workload and the resources (physical and virtual) required on those location(s).
- After checking the availability of resources in the location(s), the MCO may create them (servers, virtual machines or container clusters) if not available, or scale them, if not enough.
- The MCO deploys the workload once the necessary resources are available.

The MCO also updates and removes workloads, rescaling or releasing the corresponding resources.

For the deployment, scaling and releasing of resources, it relies on:

- the *Physical Infrastructure Manager* (PIM) for allocating the necessary physical infrastructure in certain location and
- the *Multi-Cluster Manager* (MCM) and *Virtual Infrastructure Platform manager* (VIP) to create and configure the necessary virtual infrastructure (container clusters and virtual machines, respectively) on top of it in the selected locations.

With this functionality it may handle multiple virtualisation scenarios for VNFs (components on VMs) and CNFs (cloud native components, serverless over containers, containers over VMs or over BM).

In addition, the MCO utilises the *Workload Deployment Manager* (WDM) to deploy and configure the workloads once the hosting K8s clusters or virtual machines are available, reporting back their URLs.

Wherever possible, the architecture of the MCO should reuse existing methods and functions which overlay the Kubernetes Layer (e.g. serverless implementations) and abstract deployment scenarios away from VMs or plain Kubernetes.

Physical Infrastructure Manager

The Physical Infrastructure Manager (PIM) monitors and manages a pool of physical resources (CPUs, storage, networking), selects and prepares them (with the corresponding OS and necessary software) in order to allocate these resources to a virtual machine or container cluster.

The PIM provides multiple physical infrastructure management functions, including physical resource provisioning and lifecycle management, physical resource inventory management or physical resource performance management.

The PIM follows the principles and specifications of the Physical Infrastructure Manager defined by ETSI NFV MANO [6].

Multi-Cluster Manager

The Multi-Cluster Manager (MCM) creates and configures *container clusters* both over bare metal and over virtual machines after a request from the MCO, offering a single interface to manage infrastructure from *multiple providers* and with *multiple K8s distributions*.

The MCM provides the connectors/APIs to interact with the resources and K8s distributions offered by different providers (private & public) for cluster creation, configuration and monitoring and keeps track of their evolution. The functionality provided by the MCM can be understood as the set of capabilities described by a MANO-CCM virtualisation orchestrator defined by ETSI NFV (Rel4) but with a multi-technology approach.

The MCM may create a K8s cluster on bare metal (cluster nodes are servers) or on virtualisation stack (cluster nodes are VMs), interacting with PIM or IPM respectively.

Virtual Infrastructure Platform Manager

The Virtual Infrastructure Platform Manager (VIP) creates *virtual machine* clusters across several locations using the resources allocated by the PIM.

The VIP is required when the service component to be deployed is a VAF or a CAF, which run over container clusters that make use of VMs (virtual machines).

This component works on infrastructure and technology from different providers, enabling the Telco Cloud to run on a diverse set of different virtualisation solutions (VIMs and CISMs).

Workload Deployment Manager

The Workload Deployment Manager (WDM) deploys software image(s) on top of an existing cluster(s) following MCO requests. It exposes a single interface to deploy software images (i.e. via a helm chart) on any K8s cluster based on any distribution.

The WDM provides the connectors/APIs to interact with existing clusters in different locations & technologies (K8s distributions) for NFs and third-party applications deployment and lifecycle management. The functionality provided by this component can be understood as a subset of the capabilities described by a MANO-NFVO virtualisation orchestrator as defined by ETSI NFV.

This component also deploys software images on top of virtualisation stacks (IaaS).

Virtualisation Layer (IaaS, CaaS, BMaaS)

This layer provides an abstraction of the pool of physical resources (CPUs, Storage, Networking, GPUs, NICs) to simplify its management and increase its utilisation. In this layer we find platforms that provide bare metal, virtual machines, container clusters and serverless-as-a-service (BMaaS, IaaS and CaaS).

Telco Connectivity Manager

The Telco Connectivity Manager (TCM) implements and modifies the service function chain, or removes it, totally or partially, following the requests from the TCSO (or any other system or GUI), to guarantee: a) the connectivity between workloads that will enable the service delivery and b) the connectivity from the service user to the workloads implementing the service front-end.

SECTION 2: ARCHITECTURE COMPONENTS

2.1 Multi-Cloud Orchestrator

The Multi-Cloud Orchestrator (MCO) assesses the infrastructure needed to enable a service or application, makes the infrastructure available and orchestrates actions to deploy the software for that service/application. It should be noted that workloads may be deployed on a variety of infrastructure layers: private, public or hybrid.

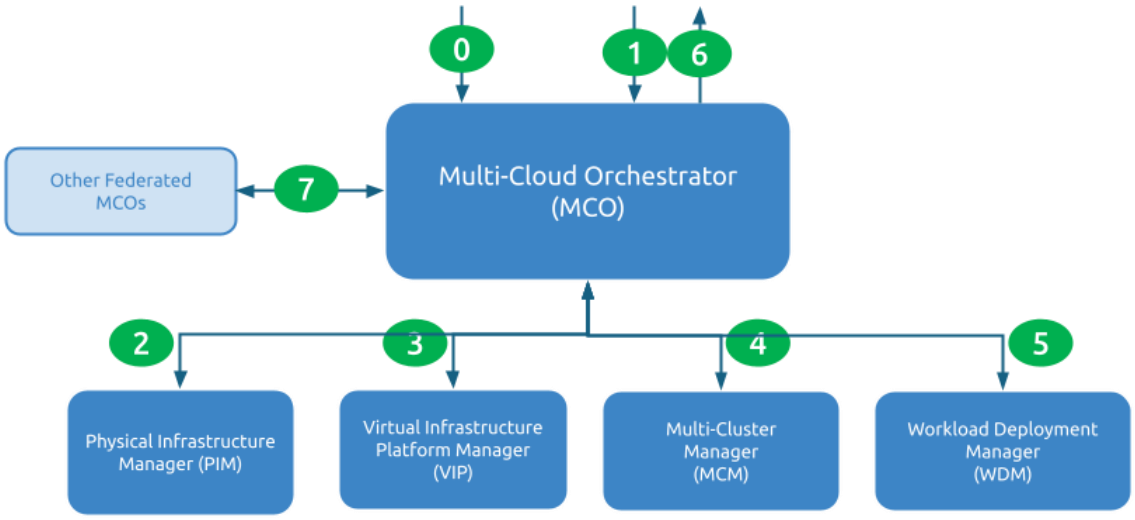

Figure 2: Interaction of the MCO with other components of the TCRA

The MCO interacts northbound with any OSS/BSS, like the *Telco Cloud Service Orchestrator*, and receives requests *(1)* that describe the workload to be deployed, the expected time of deployment (immediate or deferred) and the attributes of the workload.

The attributes of the workload include the resources required to host it, the configuration required for the infrastructure and the software package that implements the workload to be deployed.

Following the numbering in Figure 2 above, the MCO, based on the retrieved request *(1)*, identifies and selects the most suitable site(s) to deploy the workload, retrieves the software image to be deployed and manages the interaction with the other components to perform the actions to implement the decisions taken:

- Requests the MCM *(4)* to setup and configure a K8s (or requests the VIP *(3)* a VM) cluster in the selected site(s).
- Triggers the WDM to deploy the software image in the cluster that has been set up and configured *(5)*.

- Relays to the requesting system the success of the deployment and the details of the connectivity, e.g. IP address or URL to connect to the software deployed *(6)*.

The site selection may be based on the functional and technical requirements of each application (e.g., latency expected, compute resources expected) and different criteria, for instance the location of the user (to reduce latency), the site where more compute resources are available provided the latency is accepted (to balance the compute load) or the site where better performance has been measured. The criteria for site selection and the precedence/priorities in their application are defined by the entity managing the Telco Cloud *(0)* and consulted by the MCO using a function named *Telco Cloud Policy Manager (TCPM)*. The TCPM also supports the MCO by providing information regarding cost and power consumption, amongst other criteria, so that cost-effectiveness and sustainability factors are considered in the decision-making process of the MCO.

To perform the decision, the MCO needs to interact with a resource repository that stores information about the resources available in each site, the *Telco Cloud Resource Repository (TCRR)*.

In addition, the MCO keeps an inventory of the workloads that have been deployed, registering the clusters where they have been deployed. This is the *Telco Cloud Workload Inventory (TCWI)*.

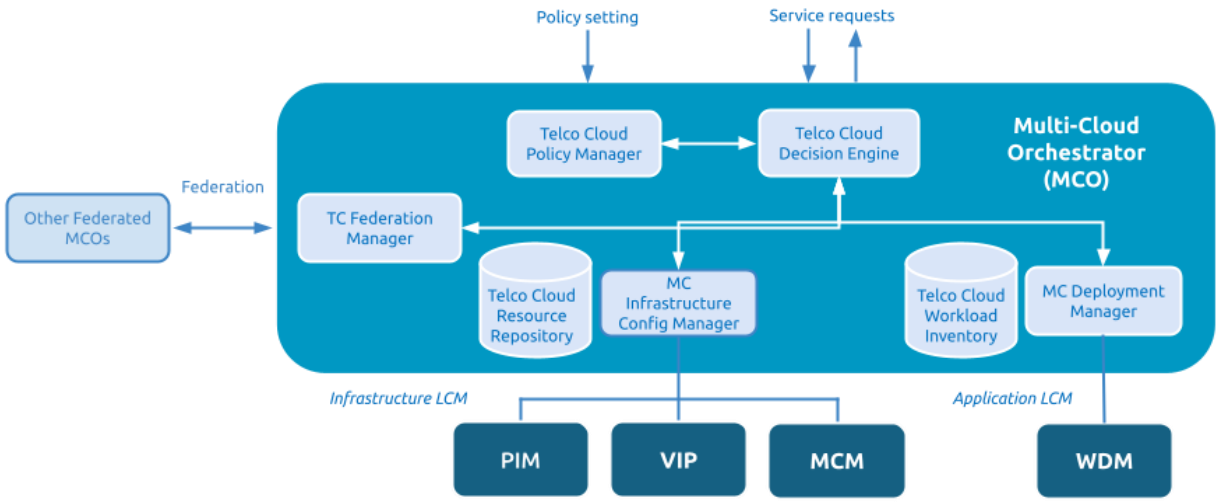

Figure 3: Internal functional components of the Multi-Cloud Orchestrator (MCO)

Setting up a cluster and/or deploying a workload may not always be required. The MCO is able to identify when these actions are required, for instance, when a K8s (or VM) cluster with spare computing resources is already available in the selected site(s) and has the right configuration for the application to be deployed, the MCO does not set up a new cluster, simply deploying the software in the existing one.

Once the software for the workload is deployed, the MCO is also in charge of guaranteeing that the infrastructure is properly configured throughout the workload lifecycle. To this end, it may interact

with the MCM (or VIP) to request specific re-configuration of an existing cluster when needed (for instance to request a change of the K8s version without impacting the workloads deployed on that cluster) or to apply external scaling of an existing cluster to make room to host an additional workload.

Note that K8s only provides an autonomous internal scaling mechanism that is limited to the number of nodes available in that cluster. An external scaling is required to increase the number of nodes for an existing cluster.

In addition, the MCO takes decisions regarding the migration of applications between sites (based on user mobility and/or performance and latency measurements from client app and network, or to balance the load of the different compute nodes) interacting with the other components to guarantee no (or minimal) service disruption while migrating.

The MCO needs updated information about the available hardware resources in the different sites to support the decisions about workload placement. To this end, the MCO contains a resource repository, the *Telco Cloud Resource Repository*, or interacts with an external one, to keep track of the computing resources that have been assigned to the deployed workloads, and of the remaining available resources, to help identify or anticipate potential resource shortages.

Federation

In cases in which different Telco Cloud Providers (TCP) are federated (i.e., have an agreement to allow roaming of Telco Cloud application users), one TCP serves applications to users of the federated TCP roaming in its Telco Cloud service area, the MCO forwards the received request to the visited operator's MCO to execute it in the visited Telco Cloud - flow *(7)* in Figure 2.

Security mechanisms like a distributed trust plane are needed to manage federation & roaming scenarios. Technologies like Self-Sovereign Identity (SSI) may facilitate the interaction between different cloud operators. In addition, when including public clouds, operators should address the aspects of data integrity and confidentiality and consider technologies such Confidential Computing.

Intelligence - AI/ML in the Telco Cloud

The following diagram depicts an overview of the components that might implement the intelligence layer required by the MCO for the decision-making processes, leading to higher levels of autonomy.

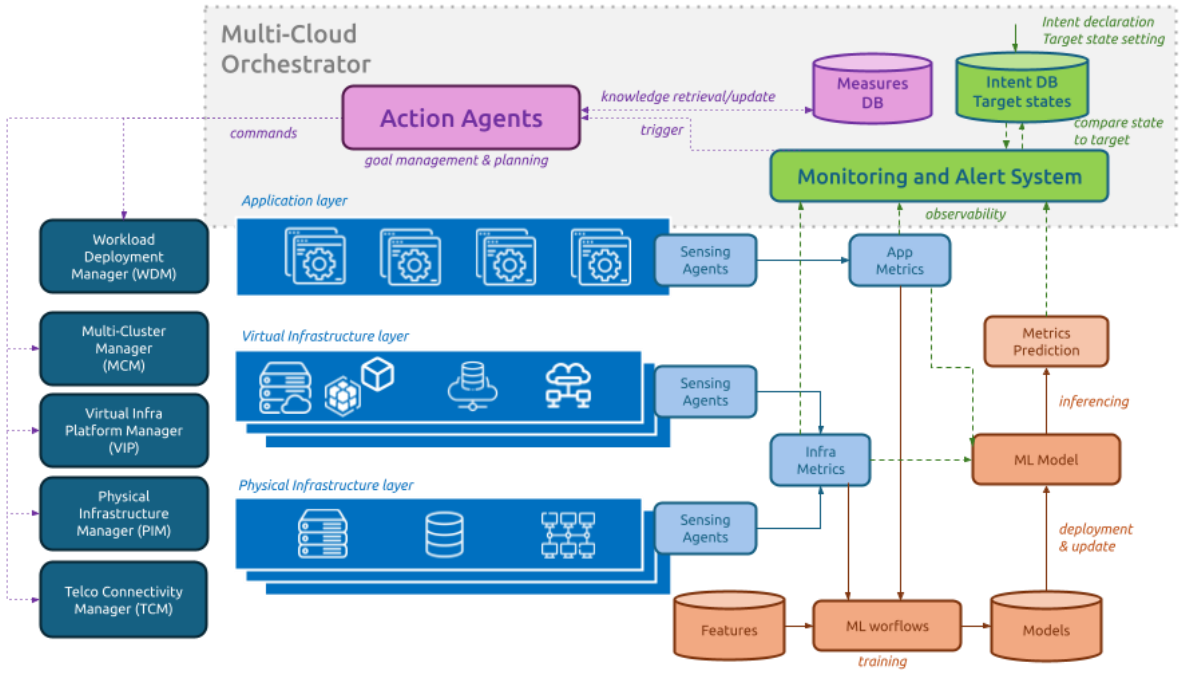

Figure 4: Elements for monitoring, learning, inferencing/prediction and action at the MCO

The intelligence system includes:

- Sensing agents that monitor the elements (compute, storage, networking, workloads/applications) at the different layers (physical infrastructure, virtual infrastructure, application) and deliver metrics that provide information about the state and evolution of the parameters describing the situation and performance of these elements.
- Training mechanisms (ML workflows) that continuously learn from these metrics and model the behaviour of the elements. These models are deployed and are regularly updated to ensure they reflect the current behaviour of the Telco Cloud. This is needed as behaviour patterns change dynamically with the evolution of the elements (infrastructure and/or applications).
- Models that are able to infer future evolution of the metrics, predicting the breach of set thresholds and thus permitting the anticipation of future deviations or problems. This may also trigger preventive actions to avoid or mitigate issues:
    - The *Monitoring and Alert System* (the observability function) receives the metrics and predictions, becoming aware of the current state and the possible future state, and checks them against the target state (goals for the different elements and systems that have been previously declared and stored in the *Intent Database*).
    - In case there is a deviation from the target state, it triggers the *Action agent*, which prepares an action plan to recover that state. To realise this, it may utilise a recommendation model based on the information of the *Measures Database*, a historical record of all measures previously taken together with the impact they have produced. The Action agent may

execute the plan directly, sending commands to the different Telco Cloud components, or ask for approval from a human operator before sending the commands.

The Monitoring and Alert System and the Action Agents are part of the MCO.

These AI/ML mechanisms also apply to the networking part, serving in the same way to the Telco Cloud Connectivity Manager (TCM).

AI/ML may be applied beyond this mere prediction or anomaly detection. The system above represents only a basic level of intelligence.

2.2. Multi-Cluster Manager

The Multi-Cluster Manager (MCM) performs the actual setup and configuration of Kubernetes clusters as requested by the MCO.

The MCM provides a single point to interact with the resources and K8s distributions offered by different providers (private & public) for cluster creation, configuration and monitoring, implementing the connectors/APIs to interface with these providers.

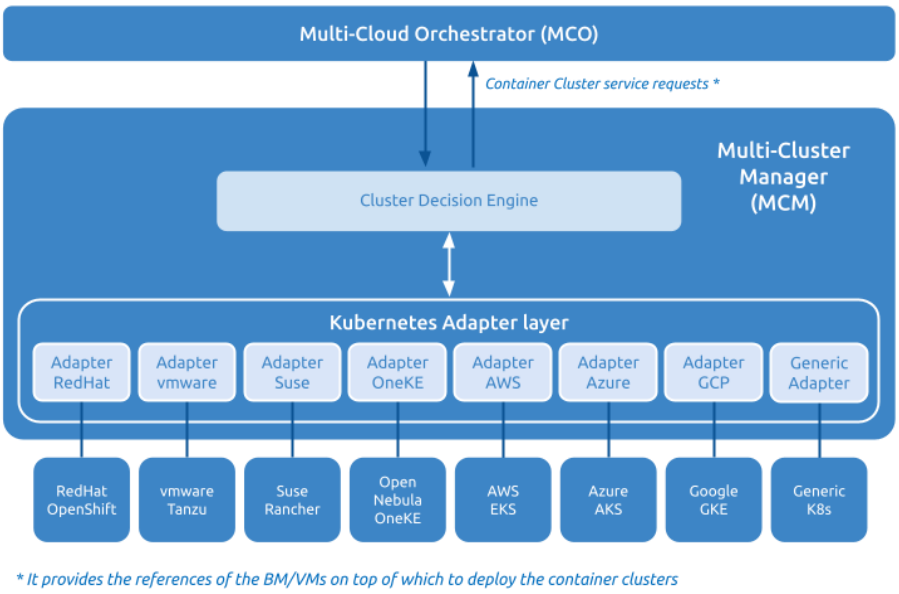

Figure 5: Internal functional components of the MCM

The MCM keeps track of the evolution of different K8s distributions and manages the connectors with different vendors exposing one single interface towards the MCO. The MCM attends to the MCO requests, setting up the infrastructure in the right K8s distributions.

This component can support the creation of K8s clusters in different modes, either on bare metal (cluster nodes are servers) or on top of a virtualisation stack (cluster nodes are virtual machines).

The MCO selects the right mode as part of the workload placement decision. Based on the numbering in Figure 6, the MCM sets up and configures the K8s cluster as requested by the MCO *(3)*, using the VMs or BM resources specified in the call *(3)*, which the MCO has previously allocated using the VIP *(2)* or PIM *(1)*, respectively.

The MCM sends the reference to the created kubernetes cluster *(5)* back to the MCO. The MCO records it in the Resource Repository and sends it to the WDM for the application lifecycle management.

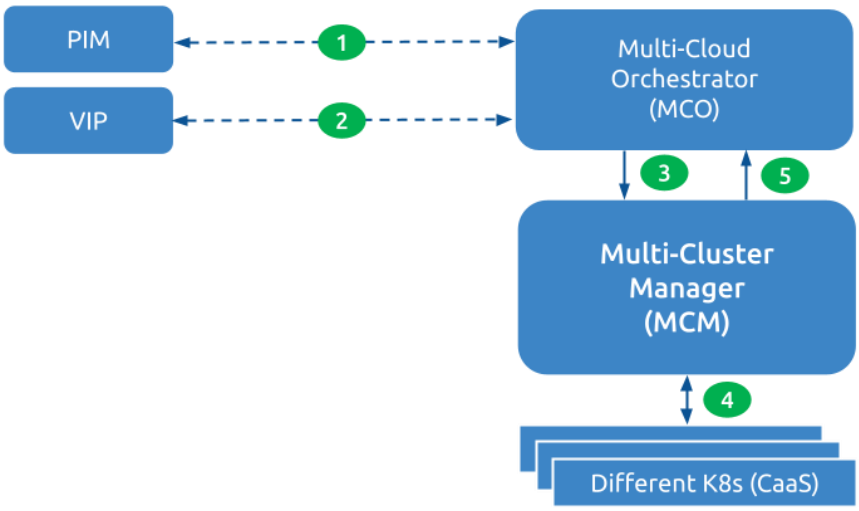

Figure 6: Interaction of the MCM with other components of the TCRA

2.3. Workload Deployment Manager

The Workload Deployment Manager (WDM) performs the deployment of software images (workloads) on top of a previously configured cluster.

The WDM provides a single point to interact with existing clusters in different locations and technologies (K8s distributions) for application deployment and lifecycle management, implementing the connectors/APIs with the different providers. The functionality provided by this entity can be understood as a subset of the capabilities described by a MANO-NFVO virtualisation orchestrator as defined by ETSI NFV.

The WDM keeps track of the evolution of different K8s distributions and manages the connectors with different vendors exposing one single interface towards the MCO. It attends the requests of the MCO, deploying the software image (i.e. via a helm chart) in the right K8s cluster.

To cope with services running on traditional virtualisation solutions, this component can also deploy software images on virtual machines (i.e. IaaS).

2.4. Virtualisation Layer

This layer is composed of functional modules of different technologies and different vendors that coexist, distributed among the different sites, delivering the traditional cloud computing services (BMaaS, IaaS, CaaS).

Typically, new services are designed as cloud-native, with software running on *containers* (CaaS), but some services are still designed following traditional virtualisation running directly over *virtual machines* (IaaS).

The decision on which virtualisation technology is to be used for a given service is taken by the service/application provider, therefore the architecture must be ready to support the different options.

The virtualisation stacks include the functionality required by the different services/applications, including among others:

- *Enhanced performance support* required to guarantee services are allocated direct access to the hardware to speed up data transfer to meet the latency required in the infrastructure (features such as Multus, DPDK, SR-IOV, CPU-pinning, and Huge pages).
- *Support for PTP* required to cope with the specific synchronisation protocol defined in O-RAN.
- *Hardware acceleration* support such as SmartNICs, GPUs, FPGA, etc. to enable cost and power-efficient implementation of demanding network functions like RAN L1 or Edge applications.
- *Support for GPU* required to cope with specific video and/or AI related edge applications.
- Support for *specific form factors* to enable virtualisation using distinct hardware that may be required for specific Edge services (e.g., lightweight servers intended for microsites requiring only frontal access).

The functionality that is supported by each virtualisation stack will be part of the information used by the MCO for the workload placement decision and should be stored in and consulted from the *Telco Cloud Resource Repository*.

2.5. Telco Connectivity Manager

The Telco Connectivity Manager (TCM) implements and modifies the service function chain, or removes it, totally or partially, following the requests from the TCSO (or any other system or GUI), to guarantee: (i) the connectivity between workloads that will enable the service delivery, and (ii) the connectivity from the service user to the workloads implementing the service front-end.

The connectivity is usually based on overlay and underlay components in each domain crossed by the traffic (e.g. WAN, data centres, etc.). The TCM manages the networking in the data centre domain through the virtualisation managers (VIM, CISM) or via specific NaaS interfaces. It manages

the WAN connectivity using Cloud Networking services (via transport SDN Controllers) for the connection of different computing nodes and uses 3GPP Core interfaces (e.g. NEF) or CAMARA APIs (e.g. QoD, Slicing, traffic influence, etc) for the user access segment.

There could be stitching points between domains to adapt the overlay and underlay technologies: for example, a *data centre gateway* could be necessary to adapt the networking between the *data centre fabric* and the Telecom Operator WAN. It may also be necessary to adapt the *data centre fabric configuration* per application, to make it consistent with the networking of the Kubernetes cluster where the application runs; and finally, it may be needed to specify the *networking translation* at the hand off point between the Telecom Operator WAN and a public cloud. Stitching points cause the need for ad hoc configurations and having many stitching points to connect different components from the user and the workload to be consumed is a barrier to automating the connectivity. The TCM manages this complexity.

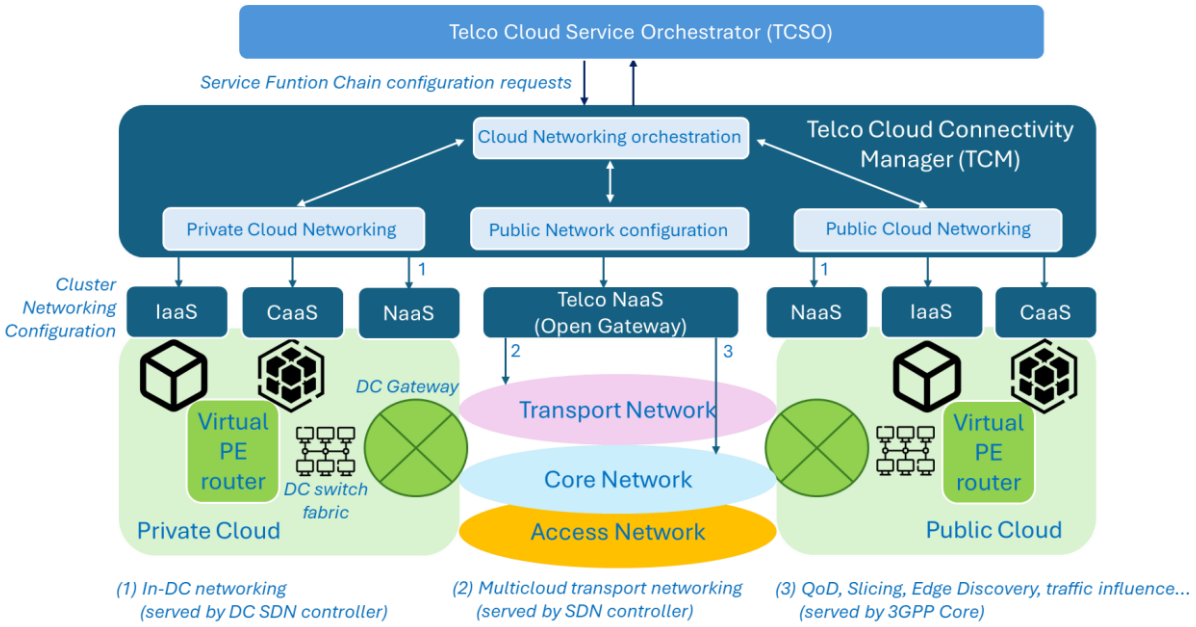

Figure 7: TCM manages the Telco Cloud networking complexity, stitching the different network segments

Adopting *networking protocol IPv6* as underlay may facilitate the stitching of virtual networks in different domains, as, unlike MPLS, it is commonly supported not only in the WAN domains but also by switching fabrics in private and public data centres. Moreover, IPv6 is even supported by servers, enabling the adoption of networking models for the Kubernetes cluster that make use of host-based routing (see below).

The overlay technology may be a choice for the domain owners, but *Segment Routing (SRv6)* [7] is the most promising one, as it natively supports *source-based routing*, *SDN connectivity control*, *networking programming* and easy migration from MPLS. SRv6 allows for *end-to-end control* of how traffic flows from the source to the destination. This reduces the reliance on routers for path

decisions and allows applications to influence the path their data takes. This is useful for applications with strict latency, reliability or security requirements.

Based on the aforementioned aims of underlay and overlay networking, an additional mechanism that is fundamental to facilitating the Telco Connectivity Manager functionalities is a mechanism to *decouple the fabric configuration from the network configuration of the Kubernetes cluster* (or from the network configuration of the virtual machine) where the application runs: this mechanism is a Host-Based Router (HBR).

An HBR refers to the use of *software-based routing on individual hosts* that implements a *virtual PE device* (i.e. Provider Edge device referring to a virtual router), responsible for managing the routing between the internal container cluster network and the external fabric network, separating the concerns of the virtualised network from the fabric. This separation allows the underlying fabric network to remain agnostic of the cluster's internal architecture and vice versa. This enables the adoption of low-cost hardware for the fabric.

An HBR allows for simpler management and orchestration. The fabric network does not need to handle the complexities of the virtual network and routing policies can be implemented per host or per application cluster. The virtual PE can therefore optimise traffic flow from the cluster to the external network, supporting different tenants in multi-tenant environments, applying appropriate routing import/export policies, firewall rules and other policies to manage communication effectively between different network domains.

Host-based routing, while flexible, tends to be slower than hardware-based routing, because it relies on the CPU and memory of the host system to process packets. Therefore, it might face scalability issues as it handles more traffic or as the number of hosts increases, as for a user plane application. Moreover, managing virtual PEs across a large number of hosts may add complexity, especially when there are many routing policies or traffic patterns to manage. These are important aspects to be addressed in future research and development.

An HBR supporting SRv6 allows path decision logic to be pushed to the edges of the connectivity (hosts). Thus, the network in between does not need to manage complex state information for every traffic flow and the end-point hosts can even specify their own routing path via SRv6 segments.

When SRv6, HBR and SDN are combined, they offer a highly *programmable and automated networking environment*. The SDN controller can dynamically optimise traffic flows in response to network events (e.g., congestion, link failures) by adjusting SRv6 segment lists.

Combining SRv6's flexibility with SDN's centralised control allows for *seamless integration between different cloud environments* (e.g., hybrid cloud, multi-cloud) with automated path provisioning.

SDN and SRv6 can *automate and optimise traffic routing* from data centres to edge locations, ensuring low-latency and high-bandwidth connectivity for time-sensitive applications.

The TCM is also in charge of configuring the different network systems to guarantee the data path to serve users is optimised for each workload. This could involve modifying the quality of service for a given user to deliver the service performance level instructed by the TCSO.

SECTION 3: EXPLORATION OF TELCO CLOUD RELATED INITIATIVES

This section identifies some initiatives that are delivering specifications or open-source implementations of components of the Telco Cloud Reference Architecture defined in this document. A mapping of the elements of each of these initiatives to the components of the TCRA is provided, demonstrating one of the values of this reference architecture mentioned in the introduction: "*provide a vendor-neutral reference definition of the Telco Cloud that… creates a level playing field for industry to discuss and compare different telco cloud initiatives and offerings*".

3.1. IPCEI-CIS

IPCEI-CIS (Important Project of Common European Interest – Next-Generation Cloud Infrastructure and Services) is a European project focused on cloud and edge computing and dedicated to developing a multi-layer interoperable and openly accessible European data processing ecosystem, the multi-provider cloud-to-edge continuum. The project [8] involves more than 100 partners aiming to advance the digital and green transition through:

- Software that will enable the necessary infrastructure-related capabilities to build the base layers of the edge-cloud stack.
- A common reference architecture to serve as a blueprint for how to set up, operate and use a cloud and edge system.
- A set of advanced cloud and edge services that can be deployed seamlessly across a network of providers.
- Sector-specific cases (for instance in the energy, health and maritime sectors) that benefit from, and show the value of, these cloud and edge services.

The research and development phase is expected to end in 2026, while the industrial deployment phase will extend until 2031 (timelines could vary depending on projects and consortia). It is expected that the first open-source reference infrastructure will be ready by the end of 2027.

This project is split into several workstreams, out of which two cover parts of the TCRA introduced in this document:

- *Workstream 1 (Cloud-Edge Continuum Infrastructure)* focuses on architecture and deployment models for the cloud-edge continuum infrastructure, searching for the best trade-off for the locations of the servers that could support the most innovative workloads. Therefore, it addresses the PIM and BMaaS while also tackling other aspects of the Telco Cloud such as the Resource Pools and the Network Infrastructure.

- *Workstream 2 (Cloud-Edge Capabilities)* is dedicated to cloud capabilities and software-defined infrastructure, developing solutions that manage the lifecycle of virtualised workloads and translate their descriptions and requirements into actions that orchestrate virtual and hardware resources, ensuring that computing capacity and networking requirements are met. Its activity spans across multiple focus areas like federation and multi-cloud orchestration, virtualisation layers (XaaS), integration, monitoring, management or optimisation. This workstream also covers cross-cutting aspects like (AI-enabled) cybersecurity, energy efficiency and interoperability. The developments driven in this workstream include some components of the Virtualisation layer (VIMs, CISMs), the Multi-Cluster Manager, the Virtual Infrastructure Platform Manager, the Workload Deployment Manager and, partly, the Telco Cloud Connectivity Manager.

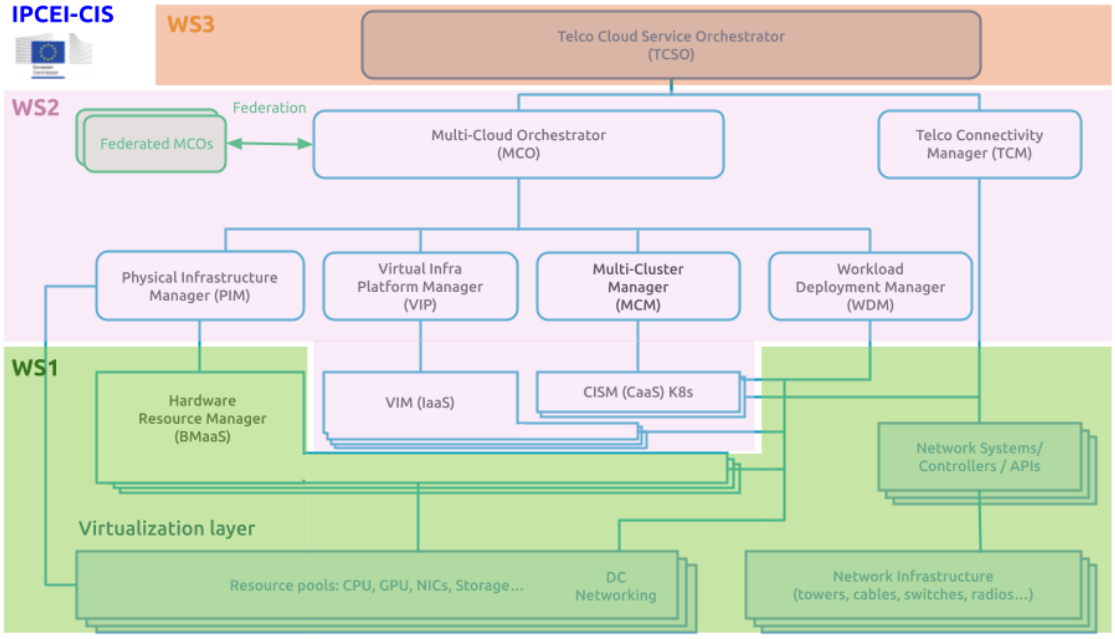

Figure 8: Mapping of IPCEI-CIS workstreams to the Telco Cloud Reference Architecture

3.2. SYLVA project (Linux Foundation)

Project Sylva is an initiative run at Linux Foundation Europe aimed at creating an open-source cloud software framework specifically designed for telecommunications and edge services. Launched in November 2022 [9], Sylva brings together leading European telecom operators and vendors, including Deutsche Telekom, Ericsson, Nokia, OpenNebula Systems, Orange, Telecom Italia, Telefónica and Vodafone.

Sylva aims to *reduce infrastructure fragmentation* by providing a unified cloud stack, to support *5G* and *edge applications* from the core to the radio access network (RAN), and to enhance *interoperability* and *simplify the deployment* of cloud-native network functions.

Sylva's first community release, Sylva V1 [10], was announced in February 2024, marking a significant milestone in the project's development. This release includes features to support 5G cloud-native network functions, manage Kubernetes clusters and improve security.

Sylva's *Telco Cloud Stack* working group is focused on releasing Telco Cloud stacks on bare metal, hypervisor (VMs) or cloud providers. These stacks compose the virtualisation layer that interacts with the Multi-Cluster Manager (MCM). The *Sylva Cluster Manager* manages the Provisioning and upgrading of Kubernetes clusters across different infrastructures (BM, public cloud, hypervisor), acting as an MCM.

Sylva's Stack, acting as a CaaS in the virtualisation layer of the Reference architecture, interacts with the Telco Connectivity Manager so that this can properly configure the routing functionality within the K8s clusters.

The Stack also includes Network Automation and an API/GUI Portal. The *Host-based routing* solution currently in technical preview aims to manage high-level automation by replacing L2 connectivity with L3 connection to data centre switches. The L3 configuration is managed in a K8s environment.

Sylva's *Workload Lifecycle Management* working group delivers best practices and tools to automate the deployment of CNFs. This automation is implemented using GitOps workflows and CNF tooling (e.g. monitoring, logging, etc) in existing Kubernetes Clusters that could be integrated in a Workload Deployment Manager.

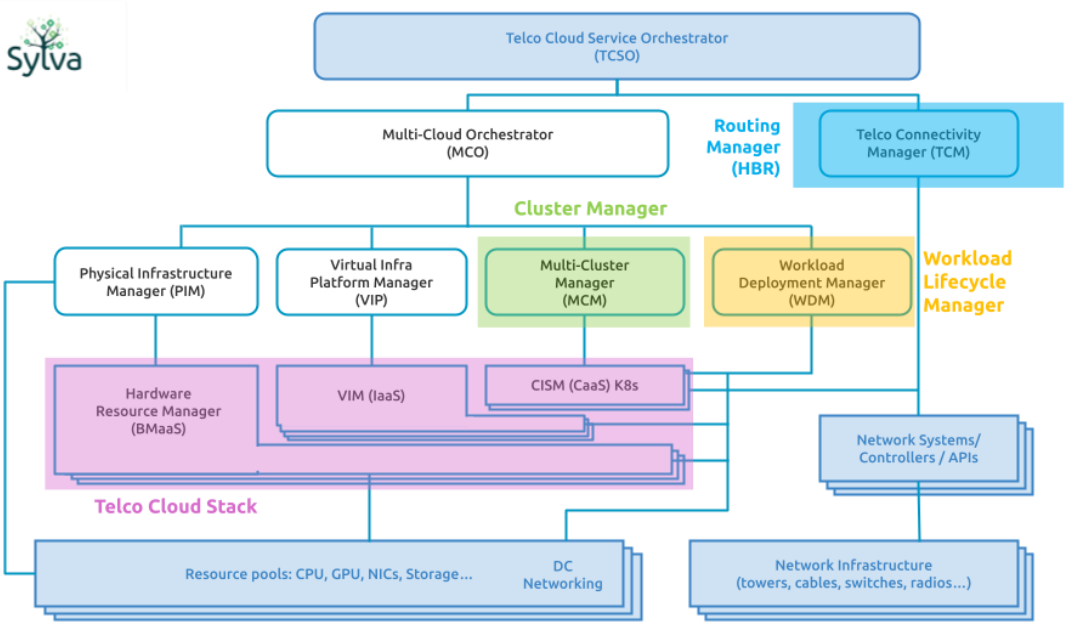

Figure 9: Mapping of Sylva components to the Telco Cloud Reference Architecture

3.3. CAMARA project (Linux Foundation)

CAMARA is an open-source project within Linux Foundation to define, develop and test Service APIs that expose Telco capabilities. CAMARA works in collaboration with the GSMA Operator Platform Group to align API requirements and publish API definitions and APIs. The harmonisation of APIs is achieved through working code created quickly and agilely with developer-friendly documentation and facilitates the availability of the APIs across telco networks and countries. This is necessary to ensure seamless customer experience, accelerate technology development and commercial adoption (minimise implementation effort), and support application portability.

Telco capabilities (supported on 4G/5G, edge, cloud and other telco systems) allow developers to get information out of the network and telco systems but also to configure them. The project provides an on-demand, secure and controlled exposure of these capabilities, paving the way for transforming operator networks into service enablement platforms, facilitating the application-to-network integration, which will be key to delivering enhanced and service-tailored customer experience.

The abstraction from Network APIs (the ones provided by the Telco systems) to Service APIs (developer-oriented) is necessary to simplify telco complexity, making APIs easy to consume for customers with no telco expertise (user-friendly APIs), to satisfy data privacy and regulatory requirements and to facilitate application to network integration.

This project is developing APIs that may be relevant as Telco Cloud service interfaces:

- The *Edge Cloud APIs* (i.e., edge-cloud resource management, application onboarding management and application instance lifecycle management) might be interfaces of a Multi-Cloud Orchestrator, Multi-Cluster Manager, Virtual Infrastructure Platform Manager or Workflow Deployment Manager.
- The *Edge Connectivity APIs* (e.g., edge discovery, traffic influence, QoD, network slicing, etc.) may be relevant interfaces for the Telco Cloud Connectivity Manager.

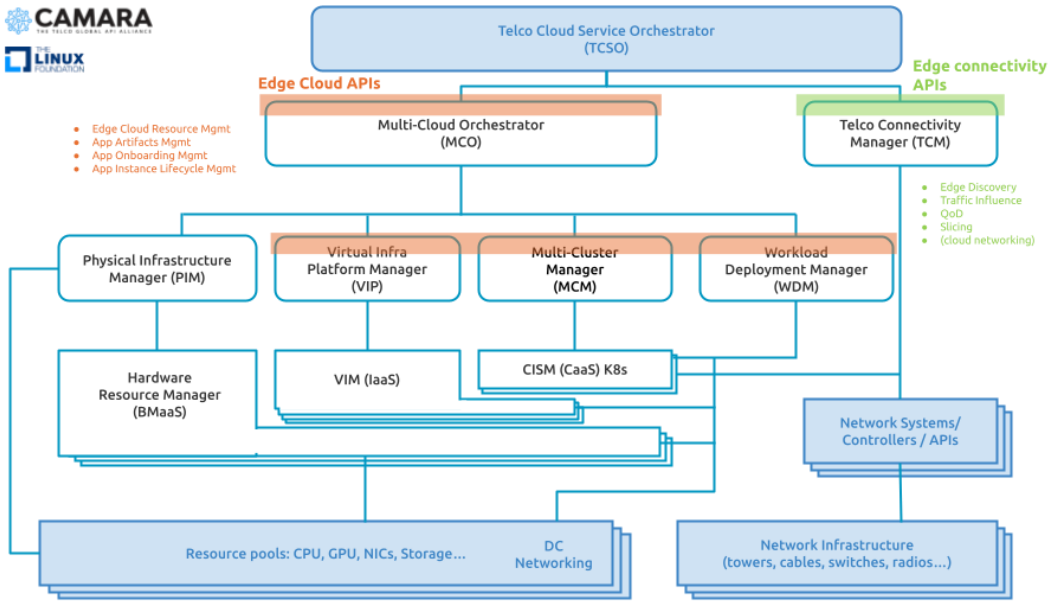

Figure 10: Telco Cloud-related interfaces covered by CAMARA project and, eventually, by GSMA Open Gateway initiative

3.4. SNS Cluster X (HEXA-X 2)

The Smart Network Services Joint Undertaking (SNS-JU or 6GSNS [11]) is an EU initiative with a mission to harness and expand European research and innovation capacities to develop advanced networks and services that will drive the digital transformation towards 2030 and beyond, aiming at positioning Europe as a global leader in digital networks and services. As part of its Strategic Research and Innovation Agenda (SRIA), it has defined its Research and Innovation Work Programme [12] (R&I WP) for 2025. One of the WP streams (Stream C) focuses on SNS Enablers and Proof of Concepts (PoCs) used to further develop and consolidate experimental infrastructure(s), in support of the various phases of the SNS JU. Stream C developments in WP 2025 have a particular focus on the 6G Telco Cloud and service platform, using open-source technologies and addressing longer term aspects of the 3C Networks orientations.

3.5. XCP-ng (Linux Foundation) and Xen Orchestra projects

XCP-ng (Linux Foundation open-source project) and Xen Orchestra (open-source project) [13] provide a full digital infrastructure management that is open, turnkey and robust. They facilitate the management, orchestration and backups of entire distributed virtual environments, delivering the following functionality:

- *Xen Orchestra:* the agentless orchestrator, telemetry, access control and backup management tool.

- *XCP-ng*: type 1, server-based hypervisor relying on the Xen technology and one of the official sub-projects of the LF's Xen project. XCP-ng is a robust and secure hypervisor that runs on a variety of x86-based hardware.

They provide the following components of the telco cloud architecture:

- The Hardware Infrastructure Manager (BMaaS) provides the ability to interact directly with physical servers and their components. Network, Compute and Storage availability, provisioning and optimisations are managed on that level. This is done directly at the Hypervisor level through *XCP-ng* and for deployment and hardware telemetry by *Xen orchestra*.
- The Physical Infrastructure manager builds on that capability and relies on the feedback of the data collected by *XCP-ng*. It is accessible through *Xen Orchestra*.
- The Virtual Infrastructure Manager handles both the virtual machines and the hypervisors running on the hosts. This is managed directly by *XCP-ng* which provides a level 1, server side and robust hypervisor.
- The Virtualisation Infrastructure Manager provides the management, observability and deployment features for virtualised environments. This is partly achieved by *Xen Orchestra*.
- The Virtualisation Layer (and specifically the CISM) component is partially supported by *Xen Orchestra* and allows for the management of containers and K8s pods enabling the lifecycle management of K8s instances.
- The Multi-Cluster Manager is equally supported by *Xen Orchestra* with the difference that it does not integrate the Kubernetes configuration for deployment but is capable of replicating or managing separate clusters and subnetworks of virtualised environments (both hypervisors and virtual machines).

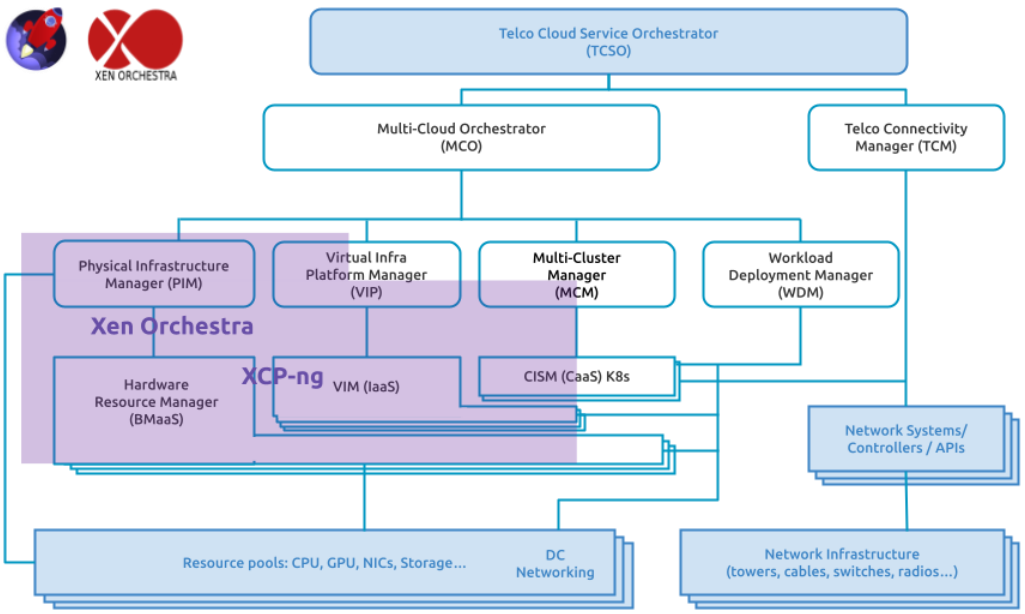

Figure 11: Mapping of XCP-ng and Xen Orchestra project components to the Telco Cloud Reference Architecture

The open-source software of XCP-ng and Xen Orchestra projects is the basis of VATES' Virtualisation Management System (VMS) [14].

3.6. OpenNebula (ONEedge5G)

OpenNebula is an open-source cloud management platform [15] delivering a simple but feature-rich and flexible solution to build and manage enterprise clouds for virtualised services, containerised applications and serverless computing. It combines virtualisation and container technologies with multi-tenancy, automatic provision and elasticity to offer on-demand applications and services.

An OpenNebula infrastructure can be deployed on-premises, in the cloud, at the edge, or in hybrid and multi-cloud environments. Virtualisation is based on the *KVM* open-source hypervisor, with support for *LXC* as well. Cloud resources are orchestrated by one or more OpenNebula front-ends (named *OneManagement*). OpenNebula can manage both single VMs and complex multi-tier services composed of several VMs that require sophisticated elasticity rules and dynamic adaptability, using the *OneFlow* component. Elements in the OpenNebula infrastructure - such as Virtual Machines, networks and appliances - are created from images and templates. OpenNebula supports the automated deployment of Kubernetes clusters through a virtual appliance, *OneKE*, the OpenNebula Kubernetes Engine.

A standard OpenNebula Cloud Architecture consists of the *Cloud Management Cluster* with the front-end node(s) and the *Cloud Infrastructure*, composed of one or several workload *Clusters* with the hypervisor nodes and the storage system, offered by *OneVirtualisation*. Infrastructure components may reside at different geographical locations and are interconnected by multiple networks for internal storage and node management. Deployment of clusters can be automated by using the OneProvision component.

OpenNebula was designed to be easily adapted to any infrastructure and easily extended with new components. The result is a modular system that can implement a variety of cloud architectures and interface with multiple data centre services.

Its capabilities are enhanced with *AI/ML* techniques and *zero-touch resource management* for dynamic deployment and operation of distributed edge environments over infrastructures with (B)5G nodes in order to optimise data processing based on criteria such as energy efficiency or latency. These enhancements have been achieved in the ONEedge5G project [16].

OpenNebula provides the following functionalities mapped on the Telco Cloud Reference Architecture:

- Physical Infrastructure Manager for allocating the necessary physical infrastructure across cloud-edge continuum being continuously aware of the availability of the resources and being

able to prepare them for allocation to VMs or container clusters. This functionality is delivered by *OneProvision*.

- Multi-Cluster Manager as it is able to create and to configure container clusters over virtual machines. Its single interface can manage infrastructure from multiple providers and with multiple K8s distributions. Furthermore, it provides the connectors/APIs to interact with the resources and K8s distributions offered by different providers (private & public) for cluster creation, configuration and monitoring and keeps track of their evolution. MCM functionality is covered by the *OneManagement* element.
- Virtual Infrastructure Platform Manager overseeing the creation of virtual machine clusters across several locations by using the resources allocated by the PIM. Also, it works on physical infrastructure from different providers. It can only create VMs based on OpenNebula virtualisation technology. For that reason, *OneManagement* covers the VIPM functionality only partially.
- Virtualisation Layer (IaaS, CaaS, BMaaS) as it provides an abstraction of the pool of physical resources (CPUs, Storage, Networking, GPUs, NICs) to simplify its management and increase its utilisation. *OneVirtualisation* delivers these services.
- *OneFlow* is the OpenNebula module that supports Workflow Deployment Manager and Telco Connectivity Manager functionality, as it allows application software images to be deployed on the clusters, and chains them to compose a service, guaranteeing the connectivity between them.
- Finally, *OneManagement* delivers part of the Multi-Cloud Orchestrator capabilities by orchestrating the allocation of virtual resources (VMs and K8s clusters) over a heterogeneous distributed computing environment.

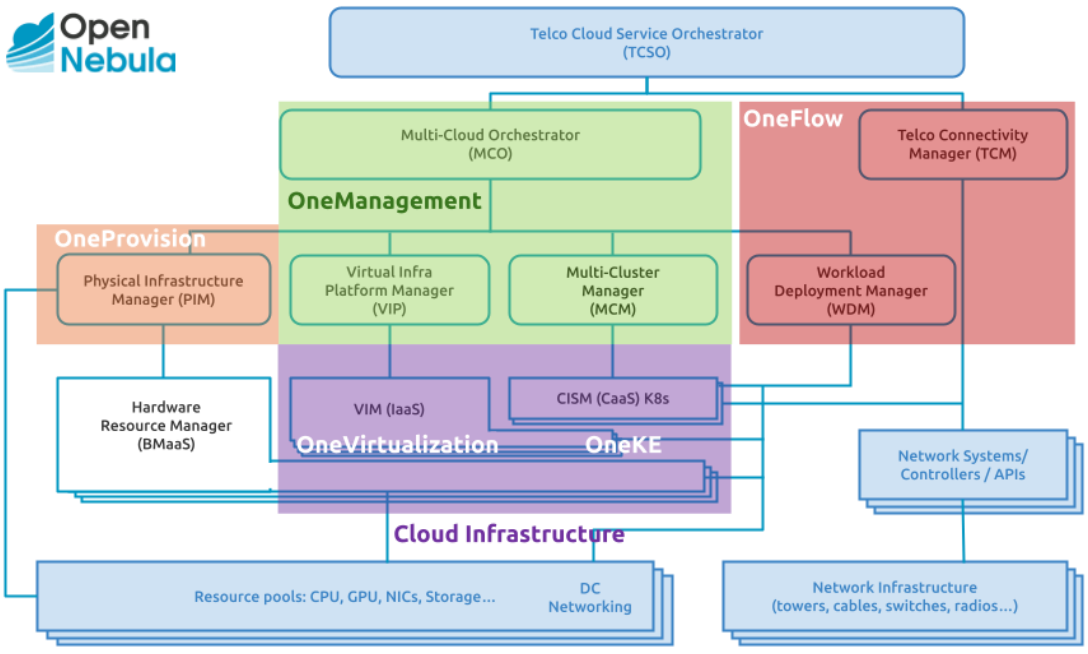

Figure 12: Mapping of OpenNebula components to the Telco Cloud Reference Architecture

OpenNebula's open-source code is freely available [17]. OpenNebula Systems [15] provides a subscription product that provides access to the Enterprise Edition, expert support, enterprise tools and knowledge base contents, maintenance packages, LTS releases and professional services.

3.7. ETSI-NFV MANO and OSM

ETSI-NFV [18] develops standards for Network Function Virtualisation, promoting cloud-native deployment and automation, and running initiatives to implement and test NFV in multi-vendor environments. ETSI-NFV is evolving towards unified network management, use of the latest cloud-native, IT, automation and AI open-source software, and multi-vendor interoperability and migration among different clouds.

NFV envisages the implementation of Network Functions (NFs) as software-only entities that run over the NFV infrastructure and identifies three main working domains:

- *Virtualised Network Functions* (both VM-based and container-based), network function software that can run over the NFVI.
- *NFV Infrastructure* (NFVI), supporting the execution of these functions, including the physical resources and their abstraction layer. It includes several components that are part of the Telco Cloud Virtualisation Layer:
    - *Virtualised Infrastructure Management* to control and manage the interaction of a VNF with computing, storage and network resources, as well as their virtualisation.
    - *Container Infrastructure Service Manager* to manage containers (e.g. K8s control plane).
    - *Physical Infrastructure Manager* to monitor and manage a pool of physical resources (CPUs, storage, networking) providing functions like physical resource provisioning and lifecycle management, physical resource inventory management or physical resource performance management.
- *NFV Management and Orchestration* (MANO), enabling orchestration and lifecycle management of resources that support the infrastructure virtualisation, and the lifecycle management of network functions. It is built on the following elements:
    - *Container Cluster Manager* (CCM) to handle container clusters over different environments (such as Kubespray, kubeadm and cluster-api). They partially address the functionality of the Multi-Cluster Manager as the CCM usually works on a single container management technology.
    - *Container Image Repository* (CIR) to enable storing of container images (like the Docker Registry) could be part of the Telco Cloud Resource Repository.
    - *VNF Manager* (VNFM), being responsible for VNF lifecycle management (e.g., instantiation, update, query, scaling, etc.), matches partially to the Workload Deployment Manager as it performs this function for workloads or cloud technologies of a certain provider (the customer needs multiple VNFMs).

- *NFV Orchestrator* (NFVO) that orchestrates and manages the NFV infrastructure and software resources, covering part of the Multi-Cloud Orchestrator functionality, and realises the network services on NFVIs across multiple domains, connecting services and NFs across multiple sites with the *WAN Infrastructure Manager* (WIM). The WIM delivers part of the functionality of a Telco Cloud Connectivity Manager.

To realise NFV in commercial products, ETSI NFV joined efforts with open-source communities (OPNFV, Anuket and now Nephio) to release NFV integration solutions based on existing IT and open-source software. Given the importance of network and service orchestration for service providers, several open-source communities such as OSM, OPEN-O and ONAP have built upon the concepts from NFV.

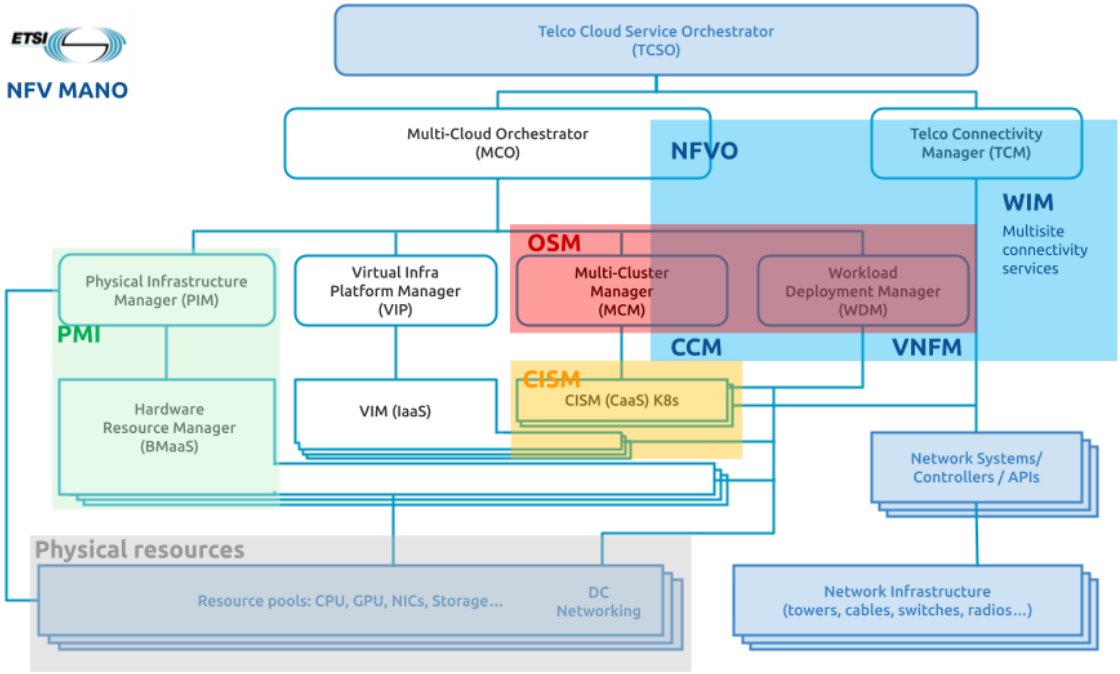

Figure 13: Mapping of ETSI-NFV (MANO, OSM) components to the Telco Cloud Reference Architecture

OSM, Open-Source MANO, is an ETSI-hosted open-source effort focused on the development of an open implementation of the NFV MANO (Management and Orchestration) stack aligned with ETSI-NFV. It is used as reference implementation and is the base for the commercial distribution of Telco Cloud orchestration platforms, and for research and demonstration of new orchestration capabilities.

OSM incorporates [19] a cloud-native approach based on a GitOps model to define and execute workflows in an intent-based declarative way so that the orchestrator keeps the requested state by means of reconciliating (i.e., synchronising) any modification of the state that diverts from the requested one.

OSM performs the following roles as defined in the Telco Cloud Reference Architecture:

- Multi-Cluster Manager, provision of a management cluster for remote cloud-native management of infrastructure. Full cloud-native management of Kubernetes clusters in public and private clouds. Azure, AWS and GCP PaaS-based clusters can be created, upgraded, scaled and deleted from OSM.
- Workload Deployment Manager, deployment of workloads in previously configured K8s clusters (CISMs) or virtual machines (VIMs). Workloads can be deployed and fully managed (upgraded, deleted) in the selected infrastructure.

3.8. Other Initiatives

The list of other initiatives and projects in communities like CNCF [20] or OpenInfra Foundation[1] [21] that address certain components of the TCRA is long. A non-exhaustive list includes: ClusterAPI, Kubermatic Kubernetes Platform, k3s, KCP, Rancher, Kubevirt, OpenStack, OVN, Metal3, Ironic. Most of them are elements of the virtualisation layer that has not been the focus of this document due to its high level of maturity and adoption.

1 https://openinfra.org/

SECTION 4: BLUEPRINTING

The TCRA has been defined to cope with complex applications. In many scenarios, a subset of the functionality of the TCRA would be enough to cover the specific requirements. This section describes an example of a blueprint to illustrate how the TCRA can be simplified. A blueprint is considered to be a reference implementation of a simplified version of the TCRA adapted to the requirements of a specific industry.

In the *manufacturing domain*, several computing tasks are performed at the edge, located at the factory floor (private edge) or close to it (far edge), due to latency, security and privacy requirements, with some applications potentially running on the cloud. In such a setting, the realisation of the TCRA will include specific components (as essential), while others would be optional, as described below:

- *Essential:* Multi-Cloud Orchestrator to manage and optimise the use of the resources at the edge, located at or close to several factory locations, and at the public cloud.
- *Essential:* Workload Deployment Manager to optimise the lifecycle management of the operational technology (OT) workloads over the distributed infrastructure.
- *Optional:* It is unlikely that in this scenario PIMs and VIPs will be needed as a single solution for virtual machines (VIM) or bare metal management (i.e. HRM) may be sufficient across the different factory locations. The multi-provider management layer makes less sense in general.

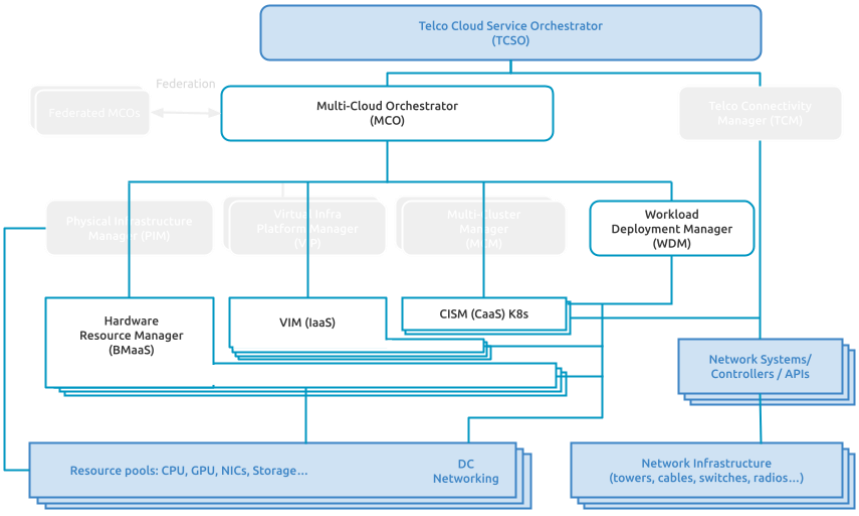

Figure 14: Blueprint for the manufacturing segment as a simplification of the TCRA

- *Optional:* The functionality required from the TCM is limited as most of the connectivity is static and can be set manually or via existing networking tools. The limited number of nodes and links makes a manual or semi-manual networking management feasible.
- *Optional:* It will not be necessary to implement a federation manager as most likely there will not be several cloud providers collaborating, or a universal service to implement.

The biggest challenge in the manufacturing scenario will not be managing the computing infrastructure and services, which are relatively static, but migrating the OT applications to a cloud-native implementation.

SECTION 5: CONCLUSIONS

This document provides a high-level overview of a Telco Cloud reference architecture that meets the requirements of complex applications, such as cloud-native telecommunication services. The complexity of a service or applications depends on the number of application component instances, the number of nodes it is distributed over (and the connectivity required between the distributed components), the amount of application component providers and the diversity of cloud technologies and services it will have to support. The aforementioned telecommunication services typically involve thousands of distributed application instances running across hundreds of nodes, interconnected in a highly dynamic environment and provided by many different providers each developing and integrating their applications on different cloud technologies and configurations.

The architecture accommodates the integration of multiple virtualisation and cloud technologies and the deployment of complex topologies of application components across a highly distributed computing environment.

Future work on the presented architecture includes:

- Integration architecture: how the different components will integrate with each other, which APIs and integration mechanisms are to be used.
- Exposure architecture: how the functionality provided by the TCRA will be exposed to customers, i.e., how it will be used by a human operator from a Telco Cloud Service Orchestrator or any other external system or application, for instance, the Service APIs that will provide access to its functionality from external applications.
- Streamlining or adaptation to less complex scenarios: not all components or functionalities of some elements within the architecture will be necessary for the applications and services required by most vertical industries. To align with the specific requirements of each sector, a simplified version of the TCRA (a vertical blueprint) will be needed to appropriately tailor and right-size it. An example was introduced in Section 4, but a more detailed blueprint definition will be required for the different vertical sectors.

Given the significant integration efforts required in the IPCEI-CIS initiative, where over 100 partners are collaborating to establish a multi-provider cloud-edge continuum, it is highly recommended that this additional architectural work should be realised within the framework of this initiative. The high level of motivation and commitment among participants provides a strong foundation for successfully tackling this complex task and ensuring alignment across all contributions. A well-defined integration architecture will significantly enhance the efficiency and success of the IPCEI-CIS project, ensuring the seamless delivery of the multi-provider edge-cloud continuum. Additionally, the development of simplified cloud blueprints tailored to various sectors will

stimulate the adoption and consumption of the edge-cloud services created within the IPCEI-CIS initiative.

Emphasising on the main gaps where concentrated development work is/will be required, focus should be on the following components:

- The Telco Cloud Connectivity Manager will be essential for implementing complex applications with highly distributed and connected components. This will apply to telecom services but also to other cases like AI-based applications with components distributed over edge devices, edge nodes and data centres. Currently, there are solutions available to manage connectivity within private data centres, public clouds and wide area networks. However, there remains a significant gap in providing seamless, end-to-end connectivity that spans across edge and cloud environments, effectively covering all segments.
- The development of Multi-Cluster Managers and Virtual Infrastructure Platform Managers will play a crucial role in enabling interoperability and portability, which are key objectives for the European Union in fostering a more open and collaborative digital ecosystem.
- Fully functional Multi-Cloud Orchestrators are currently unavailable. Dedicating efforts to their development will streamline cloud lifecycle management and optimal use of resources for Telco Cloud customers, enhancing efficiency and stimulating broader adoption of these solutions.

The TCRA should be implemented with a high level of flexibility to keep pace with the evolution of technology, and should be able to evolve to incorporate new virtualisation technology (moving beyond KVM, LXC, VMs and K8s towards others that may succeed in the future, for example, microVMs, serverless) and accommodate new hardware processing and storage technology (beyond CPUs, GPUs, DPUs, TPUs).

Security and sustainability were not addressed in the current TCRA, yet they are fundamental elements that must be incorporated into future work to ensure a robust and responsible approach.

Additionally, the potential impact of emerging trends such as WebAssembly, confidential computing, spatial computing, quantum computing, quantum cryptography or quantum networking, as well as future applications like IoT-based AI, the Metaverse, and other AI-enriched or AI-enabled applications should be explored in subsequent thematic roadmaps. This forward-looking analysis will help align the initiative with the evolving technological landscape and its challenges.

REFERENCES

[1] Cloud Alliance, https://digital-strategy.ec.europa.eu/en/policies/cloud-alliance

[2] European Industrial Technology Roadmap for the Next-generation Cloud-Edge, https://ec.europa.eu/newsroom/dae/redirection/document/102590

[3] Thematic Roadmap, "Telco Cloud: A Challenge for Next-generation Edge & Cloud", https://digital-strategy.ec.europa.eu/en/news/new-telco-cloud-thematic-roadmap-european-alliance-industrial-data-edge-and-cloud

[4] ISO Cloud Computing Reference Architecture: ISO/IEC 22123-3:2023: https://www.iso.org/standard/82759.html

[5] European Commission Digital Strategy, https://commission.europa.eu/publications/european-commission-digital-strategy_en

[6] ETSI NFV Mano Release 5: Requirements and interface specification for Physical Infrastructure Management PIM, https://www.etsi.org/deliver/etsi_gs/nfv-ifa/001_099/053/05.01.01_60/gs_nfv-ifa053v050101p.pdf

[7] IETF Segment Routing, https://www.segment-routing.net/ietf

[8] IPCEI-CIS, https://www.bmwk.de/Redaktion/EN/Artikel/Industry/ipcei-cis.html

[9] Linux Foundation Europe Announces Project Sylva, https://www.linuxfoundation.org/press/linux-foundation-europe-announces-project-sylva-to-create-open-source-telco-cloud-software-framework-to-complement-open-networking-momentum

[10] Sylva Announces First Community Release, https://linuxfoundation.eu/newsroom/sylva-announces-first-community-release

[11] Smart Networks and Services Joint Undertaking, https://smart-networks.europa.eu/

[12] https://sns-work-programme-2025_final-publication.pdf

[13] XCP-ng project: https://github.com/xcp-ng, Xen Orchestra project: https://github.com/vatesfr/xen-orchestra

[14] Vates, https://vates.tech/

[15] OpenNebula, https://opennebula.io/

[16] ONEedge5G, https://opennebula.io/innovation/oneedge5g/

[17] OpenNebula open source, https://github.com/OpenNebula/one

[18] ETSI NFV, https://www.etsi.org/technologies/nfv

[19] ETSI MANO release 16, https://osm.etsi.org/news-events/news/84-etsi-open-source-mano-announces-release-sixteen

[20] Cloud Native Computing Foundation, https://www.cncf.io/

[21] OpenInfra Foundation, https://openinfra.org/