O-RAN.WG6.TS.O2-GA&P-R004-v09.00

Technical Specification

O-RAN Work Group 6 (Cloudification and Orchestration)

O2 Interface General Aspects and Principles

O-RAN ALLIANCE e.V., Buschkauler Weg 27, 53347 Alfter, Germany
Register of Associations, Bonn VR 11238, VAT ID DE321720189

Contents

List of figures

List of tables

Foreword

This Technical Specification (TS) has been produced by WG6 of the O-RAN Alliance.

The content of the present document is subject to continuing work within O-RAN and may change following formal O-RAN approval. Should the O-RAN Alliance modify the contents of the present document, it will be re-released by O-RAN with an identifying change of version date and an increase in version number as follows:

version xx.yy.zz

where:

xx: the first digit-group is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc. (the initial approved document will have xx=01). Always 2 digits with leading zero if needed.

yy: the second digit-group is incremented when editorial only changes have been incorporated in the document. Always 2 digits with leading zero if needed.

zz: the third digit-group included only in working versions of the document indicating incremental changes during the editing process. External versions never include the third digit-group. Always 2 digits with leading zero if needed.

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the O-RAN Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in O-RAN deliverables except when used in direct citation.

1 Introduction

1.1 Scope

The present document specifies O-RAN O-Cloud functions and protocols for the O-RAN O2 interface. The document studies the functions conveyed over the interface, including management functions, procedures, operations and corresponding solutions, and identifies existing standards and industry work that can serve as a basis for O-RAN work.

1.2 References

1.2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies. In the case of a reference to a 3GPP document, a non-specific reference implicitly refers to the latest version of that document in Release 18, or the latest 3GPP release prior to Release 18 that includes that document.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, O-RAN cannot guarantee their long-term validity.

The following referenced documents are necessary for the application of the present document.

[1] Void.

[2] O-RAN.WG1.TS.OAD: “O-RAN Architecture Description”.

[3] O-RAN.WG1.OAM-Architecture: “O-RAN Operations and Maintenance Architecture”.

[4] O-RAN.WG1.O1-Interface: “O-RAN Operations and Maintenance Interface”.

[5] Void.

[6] ONAP v7.2 May 2020: “VES Event Listener Specification”.

[7] IETF RFC 6241 June 2011: “Network Configuration Protocol (NETCONF)”.

[8] IETF RFC 7950 August 2016: “The YANG 1.1 Data Modeling Language”.

[9] NIST SP 800-145 September 2011: “The NIST Definition of Cloud Computing”.

[10] 3GPP TS 29.501: “5G System; Principles and Guidelines for Services Definition; Stage 3”.

[11] ETSI GS NFV-SOL 013 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; Specification of common aspects for RESTful NFV MANO APIs".

[12] ETSI GS NFV-SOL 015 (V1.2.1): "Network Functions Virtualisation (NFV); Protocols and Data Models; Specification of Patterns and Conventions for RESTful NFV-MANO APIs".

[13] O-RAN.WG6.ORCH-USE-CASES-R004: “O-RAN Orchestration Use Cases and Requirements for O-RAN Virtualized RAN”.

[14] 3GPP TS28.550: “Management and orchestration; Performance assurance”.

[15] 3GPP TS28.531: “Management and orchestration; Provisioning”.

[16] ETSI GS NFV-SOL003 (V4.5.1): “Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; RESTful protocols specification for the Or-Vnfm Reference Point”.

[17] RFC 5424: “The Syslog Protocol”.

[18] ETSI GS NFV-SOL 009 (V4.3.1): “Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; RESTful protocols specification for the management of NFV-MANO”.

[19] ETSI GS NFV-IFA 036: “Network Functions Virtualisation (NFV) Release 4; Management and Orchestration; Requirements for service interfaces and object model for container cluster management and orchestration specification”.

[20] IETF RFC8632 ISSN: 2070-1721 September 2019: “A YANG Data Model for Alarm Management”.

[21] 3GPP TS32.111-2: “Part 2: Alarm Integration Reference Point (IRP): Information Service (IS)”.

[22] O-RAN.WG6.ASD: “O-RAN Application Service Descriptor specification”.

[23] ETSI GS NFV-SOL 001 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; NFV descriptors based on TOSCA specification".

[24] ETSI GS NFV-SOL 002 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; RESTful protocols specification for the Ve-Vnfm Reference Point".

[25] ETSI GS NFV-SOL 004 (V4.4.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; VNF Package and PNFD Archive specification".

[26] ETSI GS NFV-SOL 005 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; RESTful protocols specification for the Os-Ma-nfvo Reference Point".

[27] ETSI GS NFV-SOL 006 (V4.3.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; NFV descriptors based on YANG Specification".

[28] ETSI GS NFV-SOL 007 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; Network Service Descriptor File Structure Specification".

[29] ETSI GS NFV-SOL 014 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; YAML data model specification for descriptor-based virtualised resource management".

[30] ETSI GS NFV-SOL 018 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; Profiling specification of protocols and data model solutions for OS Container management and orchestration".

[31] ETSI GS NFV-SOL 020 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Protocols and Data Models; Specification of protocols and data models for Container Infrastructure Service Cluster Management".

[32] ETSI GR NFV-IFA 029 (V3.3.1): "Network Functions Virtualisation (NFV) Release 3; Architecture; Report on the Enhancements of the NFV architecture towards "Cloud-native" and "PaaS"".

[33] ETSI GS NFV-IFA 040 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Management and Orchestration; Requirements for service interfaces and object model for OS container management and orchestration specification".

[34] ETSI GS NFV-IFA 036 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Management and Orchestration; Requirements for service interfaces and object model for container cluster management and orchestration specification".

[35] ETSI GS NFV-IFA 005 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Management and Orchestration; Or-Vi reference point – Interface and Information Model Specification".

[36] ETSI GS NFV-IFA 006 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Management and Orchestration; Vi-Vnfm reference point – Interface and Information Model Specification".

[37] ETSI GS NFV-SEC 021 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Security; VNF Package Security Specification".

[38] ETSI GS NFV-SEC 022 (V4.5.1): "Network Functions Virtualisation (NFV) Release 4; Security; Access Token Specification for API Access".

[39] O-RAN.WG4.TS.CUS.0-R004-v17.00: "O-RAN Working Group 4 (Open Fronthaul Interfaces WG); Control, User and Synchronization Plane Specification".

1.2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies. In the case of a reference to a 3GPP document, a non-specific reference implicitly refers to the latest version of that document in Release 18, or the latest 3GPP release prior to Release 18 that includes that document.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, O-RAN cannot guarantee their long-term validity.

The following referenced documents are not necessary for the application of the present document, but they assist the user with regard to a particular subject area.

[i.1] 3GPP TR 21.905: “Vocabulary for 3GPP Specifications”.

[i.2] O-RAN.WG6.CADS: “O-RAN Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN”.

1.3 Definition of terms, symbols and abbreviations

1.3.1 Terms

For the purposes of the present document, the terms [given in [i.1] and the following] apply:

Cloudified NF: a RAN Network Function software that is deployed in the O-Cloud via one or more NF Deployments.

Deployment ID: correlation identity created by the O-Cloud for the SMO to relate to its inventory and manage.

Deployment Management Services (DMS): the DMS are the logical services provided by the O-Cloud providing for managing the life cycle of deployments which use cloud resources.

Federated O-Cloud Orchestration and Management (FOCOM): the SMO treats the collection of O-Clouds as a single federated cloud. The FOCOM are the logical services provided by the SMO to manage the distribution of O-Cloud software and provides orchestration for O-Cloud life cycle processes.

Infrastructure Management Services (IMS): the IMS are the logical services provided by the O-Cloud which provides the interface to orchestrate O-Cloud life cycle processes with the network functions it may be hosting and other operational procedures.

Managed Infrastructure Template (MIT): a representation of the declarative target of a set of managed O-Cloud Resources, which defines its input parameters and indicate the known target characteristics of the provisioned O-Cloud Resources.

Network Function Orchestration (NFO): the NFO are the logical services of the SMO which coordinates between the O-Cloud for managing deployment life cycle events, open loop, and closed loop operational procedures.

NF Deployment: a software deployment on O-Cloud resources that realizes, all or part of, a Cloudified NF.

NF Deployment Descriptor: a completed data model which provides an O-Cloud the necessary information to create a deployment.

O-Cloud: a collection of O-Cloud Resources, Resource Pools and O-Cloud Services at one or more O-Cloud Sites including the software to manage O-Cloud Resource provisioning, Nodes, Clusters and Deployments hosted on them.

O-Cloud Node: an O2ims exposed Abstracted Resource based on the Cloud infrastructure components.

NOTE: Cloud infrastructure components typically comprised of physical and/or logical CPUs, Memories, Storages, NICs, HW Accelerators, etc.

O-Cloud Node Cluster: a collection of O-Cloud Nodes that work in concert with each other, through a set of interconnecting O-Cloud Node Cluster Site Networks.

O-Cloud Platform: a platform comprising hardware and software that provides O-Cloud capabilities and services to execute RAN network functions.

1.3.2 Symbols

Void

1.3.3 Abbreviations

For the purposes of the present document, the abbreviations [given in [i.1] and the following] apply:

NOTE: An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in [i.1].

API	Application Programming Interface
FCAPS	Fault, Configuration, Accounting, Performance, Security
KPI	Key Performance Indicators
MANO	Management and Orchestration
MOC	Managed Object Class
MOI	Managed Object Instance
NIC	Network Interface Card
O-RAN	Open Radio Access Network
ONAP	Open Network Automation Platform
OSM	Open Source MANO

PM	Performance Management
PSM	Performance Subscription Manager
RAN	Radio Access Network
SMO	Service Management and Orchestration
TR	Technical Report
TS	Technical Specification

2 User defined clause(s) from here onwards

2.1 Introduction to O2

The O2 interface is an open logical interface within O-RAN architecture providing secured communication between the SMO and O-Cloud. It enables the management of O-Cloud infrastructures and the deployment life cycle management of O-RAN cloudified NFs that run on O-Cloud. The O2 interface is defined in an extensible way that enables new information or functions to be added without necessarily needing to change the protocol or the procedures. This interface enables a multi-vendor environment and is independent of specific implementations of the SMO and O-Cloud.

2.2 Alignment with O-RAN Architecture

This section provides any added detail not found in the high-level O-RAN architecture documents which define and describe the O2 interface. The intent is to extend knowledge and not to redefine existing information. Therefore, if a conflict exists between this document and the O-RAN Architecture Description [2] or the O-RAN OAM Architecture [3] the information presented will not supersede those documents.

2.2.1 Alignment with Overall Architecture

This document adopts the architecture as described in the O-RAN Architecture Description [2]. The O2 is described as the interface between the SMO and the O-Cloud to provide platform resources and workload management.

The O-Cloud needs to be created before the SMO can interface with it.

O-Cloud can consist of multiple Deployment Management Services. Each DMS can manage leased resources from multiple resource pools and span multiple locations. There is a single IMS for O-Cloud that manages all resources of DMSes as well as resources that are not allocated to any DMS in the O-Cloud. Diagram below depicts relationship between SMO and components of O-Cloud.

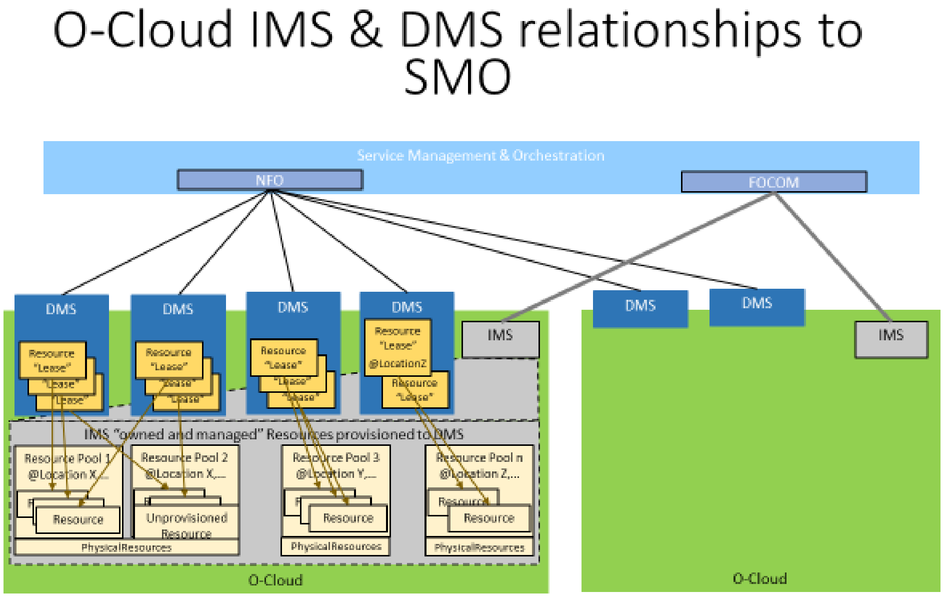

O-Cloud IMS & DMS relationships to SMO
Service Management & Orchestration
NFO
FOCOM
DMS
DMS
DMS
DMS
DMS
DMS
IMS
IMS
Resource "Lease"
"Lease"
"Lease"
Resource "Lease" @LocationZ
Resource "Lease"
IMS "owned and managed" Resources provisioned to DMS
Resource Pool 1 @Location X,...
Resource Pool 2 @Location X,...
Resource Pool 3 @Location Y,...
Resource Pool n @Location Z,...
Resource
Unprovisioned Resource
PhysicalResources
O-Cloud
O-Cloud

Figure 0-1 O-Cloud IMS and DMS Relationships to SMO

SMO can utilize blueprints or other pre-configuration templates to assign resources to resource pools and to assign resource to a DMS.

Once the O-Cloud is operational the O-RAN Architecture Description [2] describes the functions to be performed over the O2 interfaces as:

- O-Cloud Infrastructure Resource Management:
    - O-Cloud infrastructure Discovery and administration
    - O-Cloud infrastructure Scale-In, Scale-Out
    - O-Cloud infrastructure FCAPS (PM, CM, FM, Communication Surveillance)
    - O-Cloud infrastructure Platform Software Management
- Managing Abstracted Resources and DMSes:
    - Creation, Scale-In, Scale-Out of abstracted assigned O-Cloud infrastructure resources
    - Deployment FCAPS (PM, FM) for abstracted O-Cloud infrastructure resources
    - Deployment DMS (Creation, Deletion and Lease of O-Cloud infrastructure)
- NF & Services Deployment Orchestration
    - Deployment Software Management
    - Deployment, Termination, Scaling, and Healing of NF & Services deployment resources
    - FCAPS (PM, FM) for NF & Services deployment resources

These functions fall into 3 categories: managing infrastructure, managing abstract resource and DMSes, and managing NF & Services life-cycle on the DMSs.

These functions can be divided into two classes. Those which manage the infrastructure and those which manage deployments on that infrastructure. Therefore, the O-Cloud is logically composed of two functional blocks. The SMO has two logical blocks for the consumers of the O-Cloud services. The O2 as a Reference Point (RP) described in [10] are thus broken down into two (2) Service Based Interfaces (SBI) this decomposition is depicted in Figure 0-2 O2 RP Service Based-Interfaces.

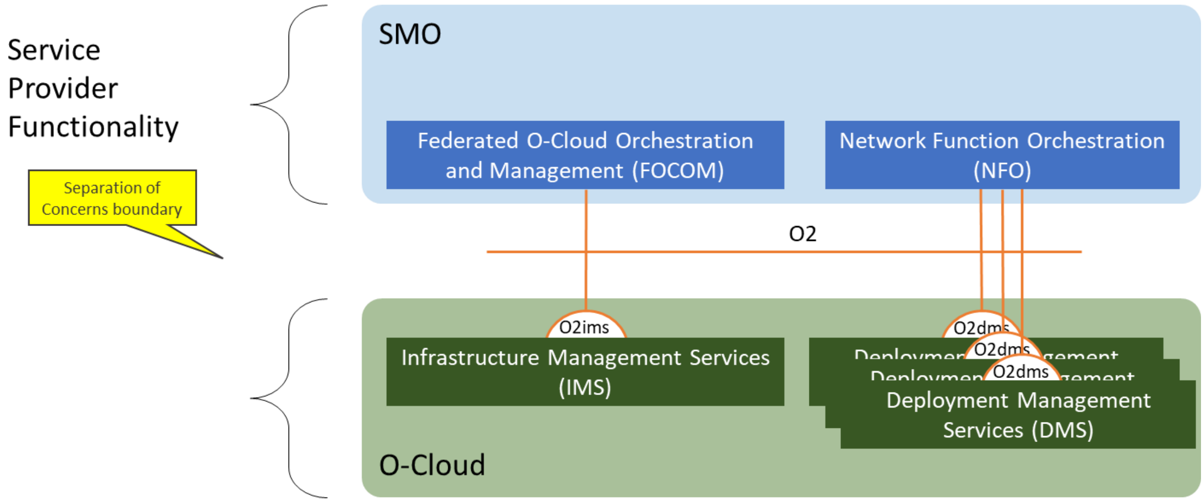

Service Provider Functionality
Separation of Concerns boundary
SMO
Federated O-Cloud Orchestration and Management (FOCOM)
Network Function Orchestration (NFO)
O2
O2ims
O2dms
O2dms
O2dms
Infrastructure Management Services (IMS)
Deployment Management Services (DMS)
O-Cloud

Figure 0-2 O2 RP Service Based-Interfaces

The following are a description of the functional blocks identified in Figure 0-2 O2 RP Service Based-Interfaces.

- **Federated O-Cloud Orchestration and Management (FOCOM)**: The FOCOM is responsible for accounting and asset management of the resources in the cloud. The FOCOM is the primary consumer of services provided by the IMS. The FOCOM has information about the O-Cloud resources management. Specifically, the FOCOM needs to know whether the services are within the operator domain or external.

- **Network Function Orchestration (NFO)**: The NFO is responsible for orchestrating the assembly of the network functions as a composition of NF Deployments in the O-Cloud. The NFO is the primary consumer of the DMS.

- **Infrastructure Management Services (IMS)**: The IMS is responsible for management of the O-Cloud resources and the software which is used to manage those resources. The IMS generally provides services for consumption by the FOCOM.
- **Deployment Management Services (DMS)**: The DMS is responsible for management of NF Deployments into the O-Cloud. It provides the ability to instantiate, monitor, and terminate NF Deployments. The DMS generally provides services for consumption by the NFO.

2.2.2 Alignment with OAM Architecture

This document adopts the O-RAN OAM architecture defined in [3], including the principles, requirements and reference architecture. It provides added detail on the use of the O2 Interface and associated functional blocks within SMO and O-Cloud.

2.3 Technology Independence

The intent of this specification is to keep it open and extensible such that it can adapt over time to technology not yet available. Therefore, functionality is described or cited with the intent to leverage open standards. There is a preference to reuse existing standards or open approaches such as Kubernetes and OpenStack.

2.4 General Requirements

The requirements below may be superseded in the O2 Specification as the interface evolves.

Table 2.4-1: O2 related general requirements

Requirement ID	Requirement	Description
REQ-O2-GEN-SMO-FUN-1	All SMOs (e.g., ONAP, OSM, etc.) shall support the O2 services and their requirements allocated to the role of the SMO.	SMO supports O2.
REQ-O2-GEN-OC-FUN-1	All O-Cloud implementations shall support the O2 services, and their requirements allocated to the role of the O-Cloud.	O-Cloud supports O2.
REQ- O2-GEN-TLS-FUN-1	Management Service providers and consumers that use TLS shall support TLS v1.2 or higher.	Communications between SMO and O-Cloud are secure.
REQ- O2-GEN-HTTP-FUN-1	Management Service providers and consumers that use HTTP shall support HTTP v1.1 or higher.	HTTP minimum is v1.1.

2.5 Relationship with Existing Standards and Open Source Solutions

The O-RAN O2 O-Cloud management services follow existing standards wherever possible. The focus of this document is to identify the use cases which conform to existing standards, identify gaps in management services for O-RAN and define needed extensions. For identified gaps, the goal is to modify the standards to include the needed O-RAN extensions and update the references in this document as the standards evolve to cover the gaps. If extensions and gaps are not specified, it is expected that the management services producers and consumers are conforming to referenced specifications.

2.5.1 ETSI NFV

This is a list of the relevant ETSI NFV standards documents which require evaluation in order to analyze packaging, LCM operations, and security for managing VMs and Containers. The versions will be specified in the O2 technical specification.

ETSI NFV standards provide a common model for the management and orchestration of VNFs (both based on VMs and containers) and NSs, as compositions of VNFs and PNFs together with network constructs for their interconnectivity.

NOTE 1: ETSI GS NFV-SOL XYZ specifications include stage 3 level specification, i.e., protocols and data models, while ETSI GS NFV-IFA XYZ specification concern stage 1 and 2 work, i.e., functional requirements, interface and information model and architecture specification.

- ETSI GS NFV-SOL004 [25] standard is used for VNF and PNF packages.
- ETSI GS NFV-SOL007 [28] standard is used for NS package.
- ETSI GS NFV-SOL001 [23] standard is used to describe VNF, PNF, and NS in TOSCA.
- ETSI GS NFV-SOL 006 [27] standard is used to describe VNF, PNF, and NS in YANG.
- ETSI GS NFV-SOL003 [16] standard is used, primarily, for VNF, LCM and Monitoring (including fault, performance and VNF indicators).
- ETSI GS NFV-SOL 002 [24] standard is used for VNF/VNFC-level EM triggered scenarios (LCM, Fault, Performance, Configuration).
- ETSI GS NFV-SOL005 [26] standard is used for NS/PNF/VNF Package Management, NS/VNF LCM and Monitoring (including fault and performance), NFVI capacity information, and VNF snapshot package management.
- ETSI GS NFV-SEC 021 [37] standard specifies specific VNF Package security requirements, while ETSI GS NFV-SEC 022 [38] standard specifies specific API and interface security mechanisms.
- ETSI GS NFV-SOL 018 [30] standard is used for the OS container workloads management provided by so-called container infrastructures services management (CISM) function that concerns to DMS. The protocol and data model of ETSI GS NFV-SOL 018 [30] profiles Kubernetes®, Helm™ and OCI™ Distribution Specification API solutions to fulfil the OS container workload, compute, storage, network, and configuration management service interface requirements specified in the ETSI GS NFV-IFA 040 [33]. These specifications are based on the study analysis performed in the ETSI GR NFV-IFA 029 [32] on enhancements of the NFV architecture towards "Cloud-native" and "PaaS".
- ETSI GS NFV-SOL 020 [31] is used for the container infrastructure services (CIS) cluster management, which concerns the management of O-Cloud Node Clusters, as equivalent abstraction in the O-RAN's defined O-Cloud domain. The protocol and data model of ETSI GS NFV-SOL 020 [31] profiles the Cluster API solution to fulfil the CIS cluster management service interface requirements specified in the ETSI GS NFV-IFA 036 [34].
- ETSI GS NFV-IFA 005/006 [35][36] covers topics Software Image Management, Virtual Resource Management (Information, Quota, Capacity, Change Notification, and Reservation) Virtual Resource Performance Management, Virtual Resource Fault Management and network forwarding path (NFP) management. ETSI GS NFV-SOL 014 [29] standard is used for the virtualised resource descriptors as input/output for the management of virtualised resources offered by the Virtualised Infrastructure Management (VIM) function that concerns to DMS.

The O2 interfaces generally follow ETSI GS NFV-SOL 013 [11] section 8 for interface security. Further refinements will be based on recommendations from the WG11.

NOTE 2: Other recommendations can be provided in future versions of the present document.

Furthermore, clauses 5.2.2, 5.3, 5.4.2, 6, 6.4, 9.1, 9.2, 9.3 and 9.4 of ETSI GS NFV-SOL 013 [11] are referenced for the relevant O2ims and O2dms RESTful APIs.

NOTE 3: Other sections of ETSI GS NFV-SOL 013 [11] are still under review and relevant recommendations can be provided in future versions of the present document.

ETSI GS NFV-SOL 015 [12] provides RESTful Patterns. Section 5.9 defines the Subscribe-Notify pattern. Generally, this pattern is adopted by the O2, with the following caveats.

- Notifications should indicate success or failure of the receipt of the notification to the notifying function.
- The subscribe and POST may not be in the same service API. The typical pattern for the O2 is that the subscribe, or its equivalent, is in the O-Cloud services and the post for the notification is an SMO service. There are no existing use cases of notifications from the SMO to the O-Cloud.
- The event types may be extended over time. Therefore, the O-Cloud services should provide the ability to discover what event types the cloud instance supports. The minimal list of supported event types shall be:

- Inventory Change
- Configuration Change
- Fault Events
- Performance Reporting
- Heartbeat

Another form of notification is the callback mechanism where the notify function is identified in the request. This mechanism is used in platform the software update procedure, as the cloud may not yet be configured for notifications when the update occurs. It is expected that all subscriptions and identified callbacks are persisted through a restart or a software update.

2.5.2 IETF

Not documented in the present document.

2.5.3 3GPP

The O2 interface is composed of services provided by the SMO and service provided by the O-Cloud. This closely resemble the Reference Points and Service Based Interface (SBI) definitions described in 3GPP TS 29.501[10]. Therefore, this specification follows the intent of the practices defined in [10] but has modified it from its intended 5G Core implementation to align with the O-RAN O2 interface.

2.5.4 Kubernetes

Not documented in the present document.

2.5.5 OpenStack

Not documented in the present document.

2.5.6 ONAP/OSM

Not documented in the present document.

3 O2 Interface General Principles

3.1 O-Clouds

O-RAN clouds are described in [i.2] as a distributed cloud composed of O-Cloud [Resource] Pools (as shown in the diagram below) where each pool is a collection of O-Cloud Resources within a Cloud Site with a specific location. An O-Cloud Node is the basic computational resource designator that is used to execute workloads and can be commonly thought of as a leased physical or logical server that is connected to the O-Cloud network. An Operator may have several O-Clouds from different vendors but will manage them as a single entity. These O-Clouds are viewed by the Service Management and Orchestration (SMO) Framework as a Federated O-Cloud.

O-Cloud Nodes are provisioned (leased) from the O-Cloud Resource Pools, with assigned Operating System and Cluster software, into managed O-Cloud Node Clusters. The O-Cloud Nodes are commonly interconnected on the Cloud Site by the O-Cloud Node Cluster Site Networks that are provisioned from the O-Cloud Site Network Fabric (where available). The workloads (e.g., micro-services or Network Functions) are deployed on the O-Cloud Deployment Plane that is constituted by the set of available O-Cloud Node Clusters with its O-Cloud Nodes and O-Cloud Node Cluster Site Networks.

O2 Infrastructure Management Services (IMS) are tailored for management and provisioning of O-Cloud Resources that are available for the SMO through the Federated O-Cloud Management (FOCOM) over O2ims. This includes allocation of the available O-Cloud Resources (e.g., Computes and Networks) into the O-Cloud Node Clusters and all cluster-wide operations on the O-Cloud Node Clusters throughout its complete life cycle.

O2 Deployment Management Services (DMS) for NF Deployment lifecycle management can be discovered over O2ims. This NF Deployment lifecycle management is available for the SMO through the Network Function Orchestration (NFO) over O2dms.

As introduced in the CADS [i.2], management of the Cloud Infrastructure where the O-Cloud is deployed is managed by the Cloud Infrastructure provider (operator internal or external). As a general concept the O-Cloud can be seen as having three functional planes that can use separate or shared resources. The control and management planes described below may be deployed on SMO-provisioned O-Cloud Nodes or outside the O-Cloud. The O-Cloud Nodes are associated with the functional plane they serve [indicated as M, C, D in the figure], to inform SMO of their roles.

- Management Plane: O-Cloud Nodes serving this plane host functions which is responsible for managing the O-Cloud, i.e., IMS and the DMS functions (indicated as M in the figure)
- Control Plane: O-Cloud Nodes serving this plane host the functions which manages the resources assigned in the deployment plane to specific deployment instances (indicated as C in the figure)
- Deployment Plane: O-Cloud Nodes and O-Cloud Node Cluster Site Networks serving this plane are used to host O-Cloud NF Deployments (indicated as D in the figure)

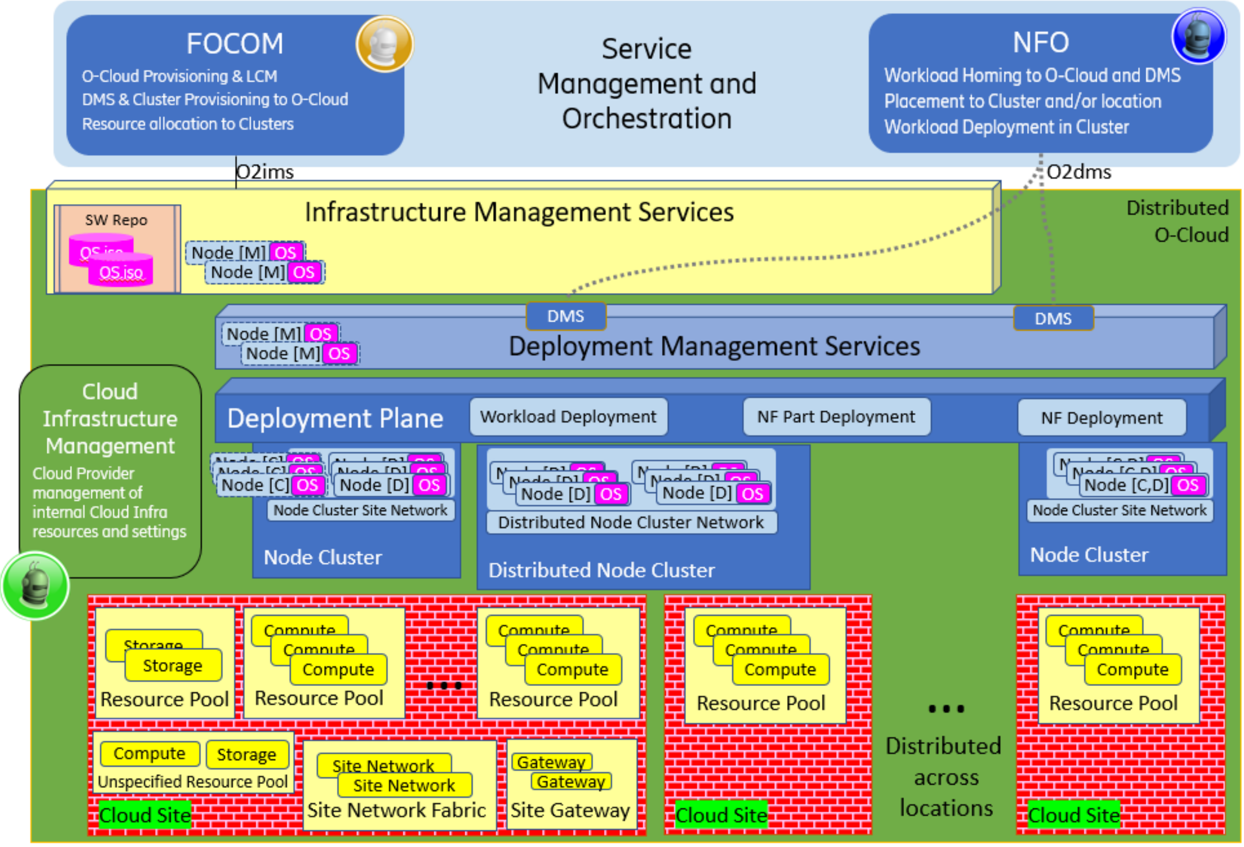


Figure 0-1 O-Cloud Services, Interfaces, and High-Level Concepts

3.2 O-Cloud Inventory

3.2.0 Overview

From the perspective of the SMO, O-Cloud Inventory consists of the exposed O-Cloud Resources based on the physical and logical Cloud Infrastructure resources, the logical clouds which it provides as interfaces for deployments, and the inventory of deployments using the logical clouds. One of the responsibilities of the SMO is for “financial” accounting of the physical and logical inventory. The SMO inquires about the mapping of resources to their use. The O-Cloud is responsible for the assignment of resources to their use. Typically, the O-Cloud inventory is the result of the SMO provisioning requests and assignments of O-Cloud Resources with additions of O-Cloud internal status of the existing internal resources, e.g., if some of the resources are down due to faults or maintenance operations.

3.2.1 Infrastructure Inventory

Infrastructure inventory is passed between the O-Cloud and the SMO over the O2ims. An O-Cloud Resource Identifier is used to correlate the inventory. This identifier needs to be discoverable by the O-Cloud and would likely be on an "invoice" for the O-Cloud "purchase order" or whatever method that is applied to account for requested resources. The O-Cloud Resource Identifier is to be immutable.

The IMS will add O-Cloud Resources to the relevant O-Cloud Resource Pool(s) when they are ready to be provisioned as an O-Cloud Resource. The IMS can then allocate the available O-Cloud Resources to O-Cloud Nodes and O-Cloud Site Networks when there are needs in the O-Cloud Node Clusters. SMO is then notified or can discover the O-Cloud Node, O-Cloud Node Cluster Site Network and O-Cloud Node Cluster.

The SMO receives O-Cloud Resources startup events and updates its inventory accordingly. The O-Cloud Resource Identifier is matched with "invoice" data in order to provide "financial" asset tracking reports or similar assurance reports. The location and use identifiers are used to track inventory which will be monitored as part of cloud service assurance.

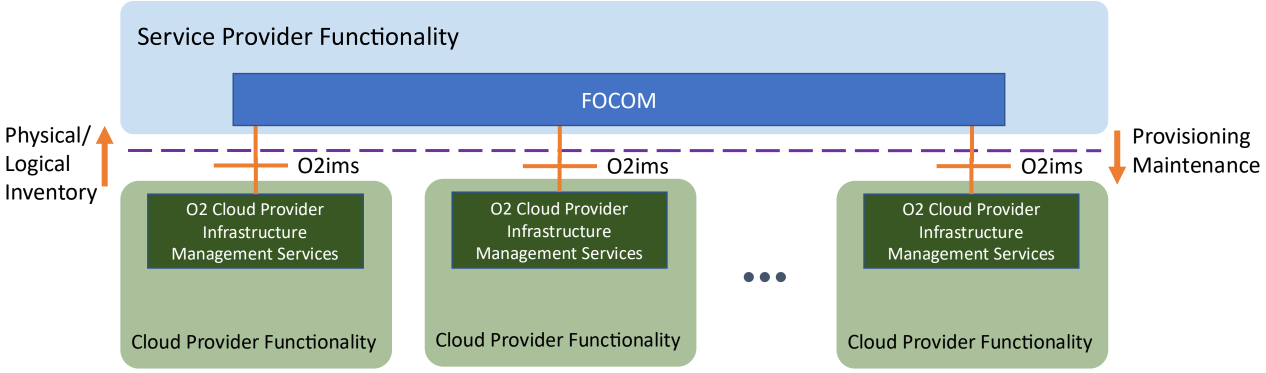

Service Provider Functionality
FOCOM
Physical/ Logical Inventory
Provisioning Maintenance
O2ims
O2ims
O2ims
O2 Cloud Provider Infrastructure Management Services
O2 Cloud Provider Infrastructure Management Services
O2 Cloud Provider Infrastructure Management Services
Cloud Provider Functionality
Cloud Provider Functionality
Cloud Provider Functionality

Figure 0-1 O-Cloud Infrastructure Inventory

3.2.2 Logical Inventory

The O-Cloud itself needs to have one or more DMS available within its distributed footprint to enable NF Deployments. These could be based on different virtual cluster technologies e.g., Kubernetes/Docker and Open Stack. Each DMS endpoint provides an O2dms interface and is inventoried by the SMO as a Logical Cloud. The Logical Cloud and its Cloud Site locations are used by the SMO in order to select the O-Cloud to be used for a deployment during the Logical Cloud selection process.

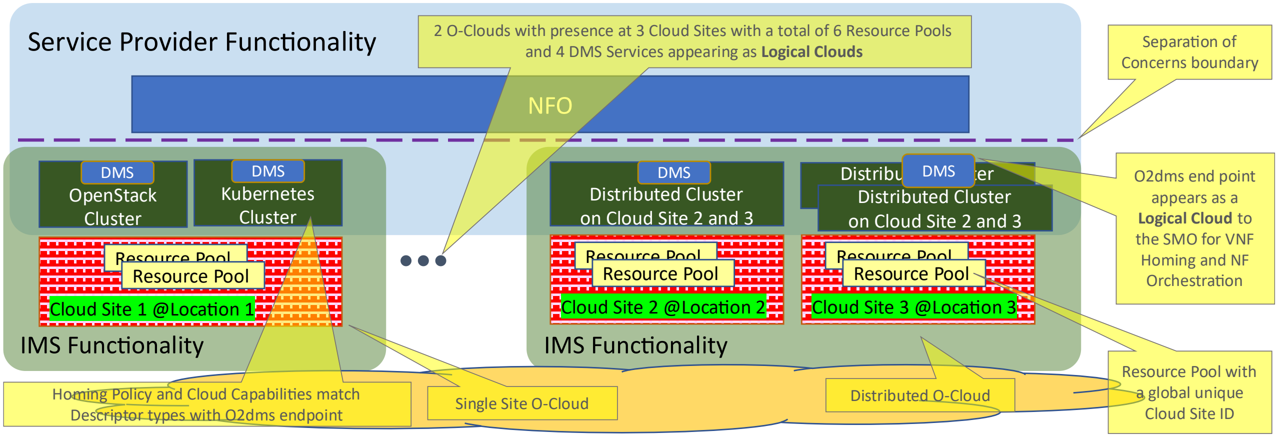

Service Provider Functionality
2 O-Clouds with presence at 3 Cloud Sites with a total of 6 Resource Pools and 4 DMS Services appearing as Logical Clouds
NFO
Separation of Concerns boundary
DMS
OpenStack Cluster
DMS
Kubernetes Cluster
Resource Pool
Resource Pool
Cloud Site 1 @Location 1
IMS Functionality
DMS
Distributed Cluster on Cloud Site 2 and 3
Distribu DMS ter
Distributed Cluster on Cloud Site 2 and 3
Resource Pool
Resource Pool
Cloud Site 2 @Location 2
Resource Pool
Resource Pool
Cloud Site 3 @Location 3
IMS Functionality
O2dms end point appears as a Logical Cloud to the SMO for VNF Homing and NF Orchestration
Resource Pool with a global unique Cloud Site ID
Homing Policy and Cloud Capabilities match Descriptor types with O2dms endpoint
Single Site O-Cloud
Distributed O-Cloud

Figure 0-2 Logical Clouds Example

3.2.3 Deployment Inventory

Although the cloud could support IaaS and PaaS deployments, most O-RAN MEs are composed of SaaS deployments as defined by NIST [9]. The Deployment Descriptor is a set of metadata contained in a datatype matching a capability advertised by the O-Cloud. Logical Cloud selection in the NFO will match the descriptor type with the capability type to enable deployment. The Deployment Descriptor will cause one or more cloud resources to be allocated (VMs, Pods, Containers, Networks) but will return a single correlation ID, referred to here as the Deployment ID. The O-Cloud will provide, over the O2dms, the ability to fetch the details of these objects from their Deployment ID.

3.3 O-Cloud Monitoring Service

3.3.0 Overview

When the O-Cloud Infrastructure or the ORAN cloudified NFs fails, it needs to be fixed immediately, and preferably automatically, to prevent end users from experiencing service disruptions. To avoid this service disruption Network Operations must consider the telemetry information of O-Cloud deployments in the network. The telemetry information serves as a vital resource for analysing the O-Cloud's state and health, and for delivering on service monitoring goals.

The O-Cloud Monitoring Service uses telemetry data to provide monitoring of O-Cloud infrastructures. The telemetry data is provided by Deployments and Infrastructure and is correlated to reduce duplication of effort in determining the root cause in the network.

There are different types of telemetry, Managed Element Telemetry, Deployment Telemetry, and Infrastructure Telemetry.

1. For Managed Element Telemetry, the objective is to monitor the application behaviour. This will be addressed in the O1 Specification [4].

2. For Deployment Telemetry, the objective is to monitor the number of deployment instances an O-Cloud has at that moment and how many were expected, how the on-progress deployment is going, and health checks. Additional Deployment Telemetry metrics like CPU, network, and memory usage can also be collected. This will be addressed in the O2 Specification.

3. For Infrastructure Telemetry, the objective is to monitor the health of the O-Cloud Infrastructure components. Network Operations are interested in discovering if all the components in the O-Cloud Infrastructure are working properly and at what capacity, how many deployments are running on each node, and the resource utilization of the O-Cloud Infrastructure. This will be addressed in the O2 Specification.

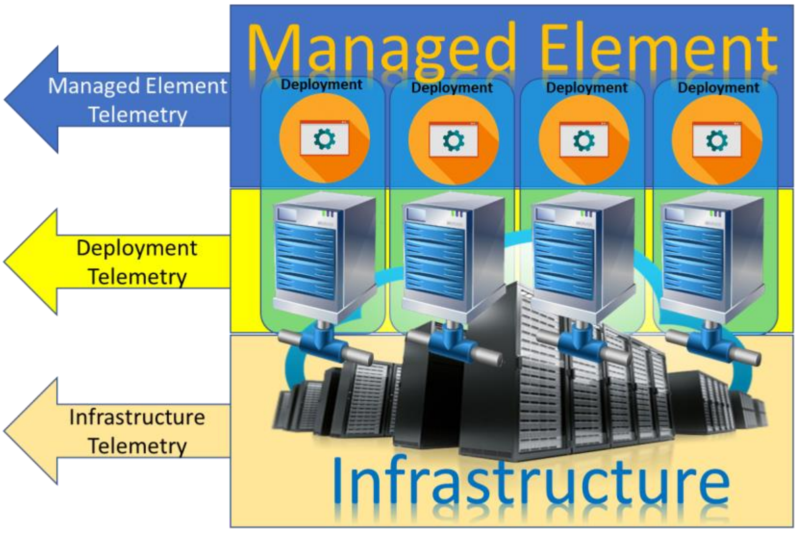


Figure 0-1 Telemetry Types

3.3.1 General Capabilities

Business Capabilities:

- As Network Operations I need telemetry data from Network Functions in order to provide service assurance of the network (Note: This is assumed to be handled by the O1)

- As Network Operations I need telemetry data from the infrastructure in order to provide service assurance of the network

- As Network Operations I need to be able to correlate Network Function Telemetry and Infrastructure Telemetry in order to reduce duplication of effort in determining one root cause

Functional Capabilities:

- The SMO shall be able to correlate ME telemetry to Infrastructure and Deployment telemetry to aggregate problems to a root cause
- The O-Cloud shall be able to make all Configuration Data and any external changes to it available to the SMO
- O-Cloud telemetry shall minimally consist of Fault, Performance, and Configuration Data
- The SMO shall be able to correlate a Managed Element to its deployment components
- The O-Cloud shall be able to report telemetry of deployment resources relative to those identified in the deployment descriptor
- The O-Cloud shall be able to report Infrastructure telemetry and identify the deployments using the resource
- O-Cloud shall provide the collection and reporting of performance information of O-Cloud resources
- O-Cloud shall support the capability to notify about the availability of performance information
- O-Cloud shall expose the type of performance information that can be collected for the allocated O-Cloud resource(s)
- O-Cloud shall expose the type of O-Cloud resource, for which the performance information can be collected
- O-Cloud shall provide the collection of fault information for O-Cloud resources
- O-Cloud shall support providing notification of fault information related to O-Cloud resources

3.4 O-Cloud Provisioning

3.4.0 Overview

O-Cloud Provisioning is the allocation and configuration of O-Cloud Resources and Services to an O-Cloud Node, Node Group, Node Cluster Site Network, Node Cluster and/or Logical Cloud (as represented by a DMS). This is one of the key functionalities of the O-Cloud, relating to how an O-RAN Cloudified Network Function or part thereof can be deployed on the O-Cloud services and resources. The O-Cloud resources are deployed flexibly over the distributed O-Cloud Sites to match the O-RAN Cloudified Network Function's fluctuating demands. The O-Cloud Node Clusters and Logical Clouds typically scale out using additional O-Cloud Resources to accommodate spikes in usage and scale in by releasing the used O-Cloud Resources when demands decrease. O-Cloud Provisioning enables numerous benefits including scalability, speed, and cost saving.

To support the deployments, O-Cloud Provisioning will need to provide several functionalities. There will be initial support for the following:

- Affinity, Anti-Affinity, Quorum Diversity Rules
- Capacity Query
- Availability Query
- Managed O-Cloud Node Clusters and Logical Clouds
    - Creation, Deletion, and Scaling for Cloud Site localized Clusters
- Identification of the O-Cloud Capabilities which are relevant for NF Deployments, that realize, all or part of, a Cloudified NF, as defined in Annex D

Support for the following functionalities is not addressed by the present version of the specification:

- Subscriptions
- Flavor Management

- Server Group Management
- Capabilities Discovery
- Switch Fabric Management
- O-Cloud Gateway Management
- Managed O-Cloud Node Clusters and Logical Clouds for distributed Clusters over multiple Cloud Sites
- Quota Management
- Resource Performance Management
- Virtual Resource Performance Management

3.4.1 Affinity, Anti-Affinity, Quorum Diversity Rules

3.4.1.0 Overview

An Affinity rule specifies that deployments with the same rule applied must be collocated within the same scope. This can be applied to a single O-Cloud Node or an O-Cloud Node Group. Anti-Affinity is the opposite of the Affinity Rule in that deployments with the same rule applied cannot be collocated within the same scope. Affinity and Anti-Affinity rules can be applicable to entities outside of the current deployment such as Tenant or Namespace.

A Quorum diversity rule stipulates that deployments with this rule can be collocated if less than 50% of the deployments exist within the same scope. This rule can only be enforced when there is a minimum of 3 deployments. In which it will be the same deployment assignment as Anti-Affinity. In scenarios with more than 3 deployments a minimum of 3 different resource assignments at the specified scope (O-Cloud Nodes or O-Cloud Node Group) will be used. More can be used at the discretion of the O-Cloud Resource Scheduler.

3.4.1.1 General Capabilities

Business Capabilities:

- Network Function Orchestration (NFO) shall be able to apply Affinity, Anti-Affinity, Quorum Diversity rules over the O2dms interface

Functional Capabilities:

- O-Cloud Provisioning shall provide Create rule for Affinity, Anti-Affinity, and Quorum Diversity
- O-Cloud Provisioning shall provide Read rule for Affinity, Anti-Affinity, and Quorum Diversity
- O-Cloud Provisioning shall provide Update rule for Affinity, Anti-Affinity, and Quorum Diversity
- O-Cloud Provisioning shall provide Delete rule for Affinity, Anti-Affinity, and Quorum Diversity
- O-Cloud Provisioning shall provide NFO with means to perform NF Deployments according to rules for Affinity, Anti-Affinity, and Quorum Diversity

3.4.2 O-Cloud Capacity and Availability

3.4.2.0 Overview

Cloud Infrastructure Providers install and configure the available O-Cloud Resources for an O-Cloud. O-Cloud Resources are allocated to O-Cloud Node Clusters and configured by the IMS to enable NF Deployments by the DMS where each DMS may have a different distribution of Capacity, and Availability.

The O-Cloud Capacity is the number of exposed O-Cloud Resources of each ResourceType that can be allocated or reserved for O-Cloud Node Clusters. The O-Cloud Availability is the currently free number of exposed O-Cloud Resources of each ResourceType that is not allocated or reserved. The Capacity may be larger than the sum of available exposed physical resources if the O-Cloud allows oversubscription.

3.4.2.1 General Capabilities

Business Capabilities:

- Network Operations shall be able to view the Capacity and Availability of an O-Cloud

Functional Capabilities:

- O-Cloud Provisioning shall provide Query of O-Cloud Capacity
- O-Cloud Provisioning shall provide Query of O-Cloud Capacity allocations and reservations
- O-Cloud Provisioning shall provide Query of O-Cloud Availability

3.4.3 O-Cloud Provisioning Basic Concepts

3.4.3.0 Overview

These basic concepts will become relevant moving from the traditional RAN paradigm of total network and resource ownership to cloudified networks where there are more stakeholders, and the resources are shared amongst stakeholders. This section will describe basic concepts related to provisioning in the context of IMS and DMS. This section will elucidate that the IMS and DMS have a difference in kind. Provisioning is basic OA&M functionality. The concepts described here will motivate the development of O-Cloud O2 IMS & DMS provisioning service modelling, and information modelling.

Provisioning in the context of the GA&P is the act of providing access to entities, e.g., resources.

The general concepts described here relate to O-Cloud entities as well as provisioning functionality. The subsections group together like concepts.

3.4.3.1 Workload

A workload is any software application(s) that consumes an Abstracted Resource (as defined in this document).

For example, this workload could be a Host Operating System, Guest Operating System, micro-service, or the software part of a Physical Network Function (PNF), Virtualized Network Function (VNF) or Cloud Native Network Function (CNF) which would typically be deployed by a DMS.

3.4.3.2 Resource Related Concepts

3.4.3.2.1 Resource

An O-Cloud Resource is defined in an O-Cloud infrastructure as one that may be provisioned and managed through an abstraction. O-Cloud infrastructure has two kinds of resources: physical and logical resources. A physical resource is defined as a kind of resource that has a manifestation in the real world. A logical resource is defined as a kind of resource that uses a number of physical resources, or a portion of physical resources, or a number of software entities. A resource can be assigned to an orchestrator that will deploy workloads for resource consumption (see figure 1).

For example, some kinds of physical resources could be servers, switches, and storage units.

For example, some kinds of logical resources are Openstack and K8S clusters.

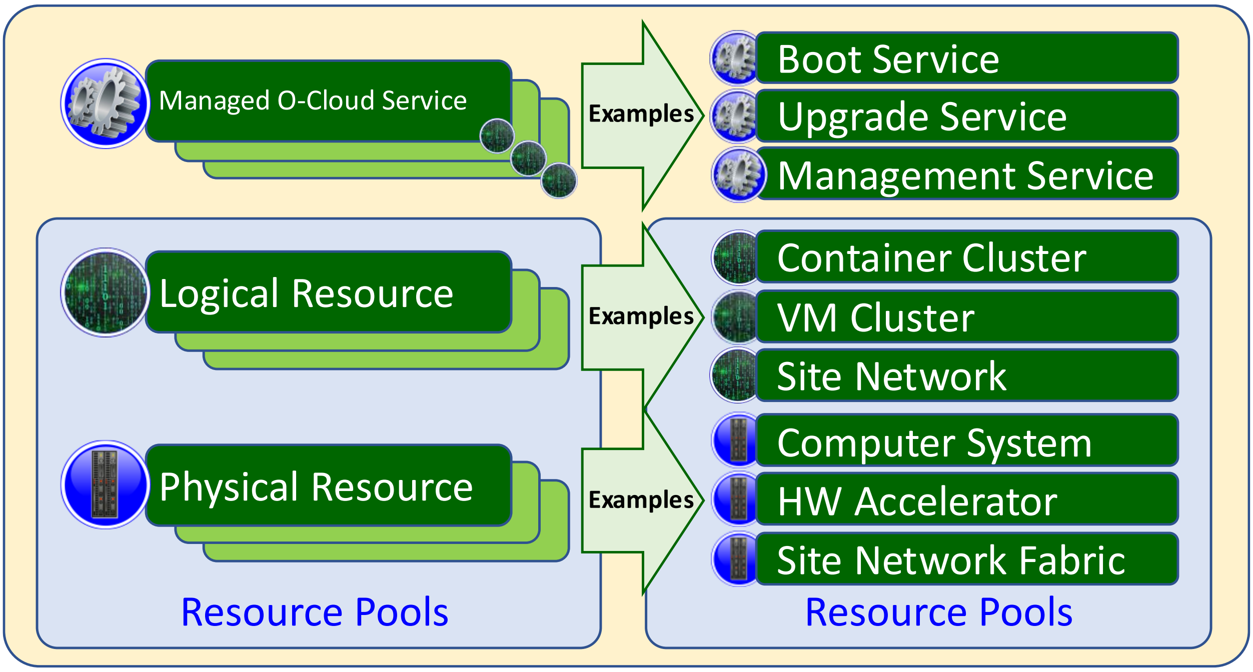


Figure 0-1 Diagram of kinds of Resources and examples of Resources

3.4.3.2.2 Managed O-Cloud Service

A Managed O-Cloud Service is an exposed, provisioned, monitorable, or managed entity that performs some operations for a client or consumer. A Managed O-Cloud service is an entity that provides some functionality to other entities. Managed O-Cloud Services could execute on physical and/or logical cloud infrastructure resources thereby consuming cloud infrastructure resources not usable by the O-Cloud. Managed O-Cloud Services could also execute outside of the cloud infrastructure resources. Unlike logical services, Managed O-Cloud Services are created on demand ad infinitum. The IMS is the management endpoint for the Managed O-Cloud Service.

NOTE: Exposure means it is visible outside of the IMS through the IMS endpoint as a managed object; thus, the SMO is aware of it.

For example, an autonomous service as a managed O-Cloud service, which as a plug-in to the cloud that performs some unique functions upon resources.

For example, some kinds of Managed O-Cloud service might be boot services, upgrade services, and DMS.

3.4.3.3 Resource Views

3.4.3.3.0 Overview

A *Resource View* is defined as the visible and/or accessible part of the Resource that an actor sees based on their authority. Resources are seen by different actors that need to manage or act upon resources. Different perspectives from these actors are "views" of the Resources. O-Clouds can have multiple Resource Views to enable separation of authority and operations of the resources, into logical groups of actors (generally other SW entities). The four kinds of O-Cloud resource views are: *Cloud Infrastructure Resource, Abstracted Resources, Assigned Resource, and Consumed Resource*. This is done to provide maximum separation of concern between different actors. These views accommodate public, external, and shared private cloud deployments.

For example, Consider a hotel room as a resource. The front desk staff who assigns a room to a tenant has a relationship to the room and job requirements of how to manage the room. That is, they have a "view" of the room. The tenant that checks into the room has a different "*resource view*" of that same room. The cleaning staff has yet another relationship, or "resource view", to the same hotel room.

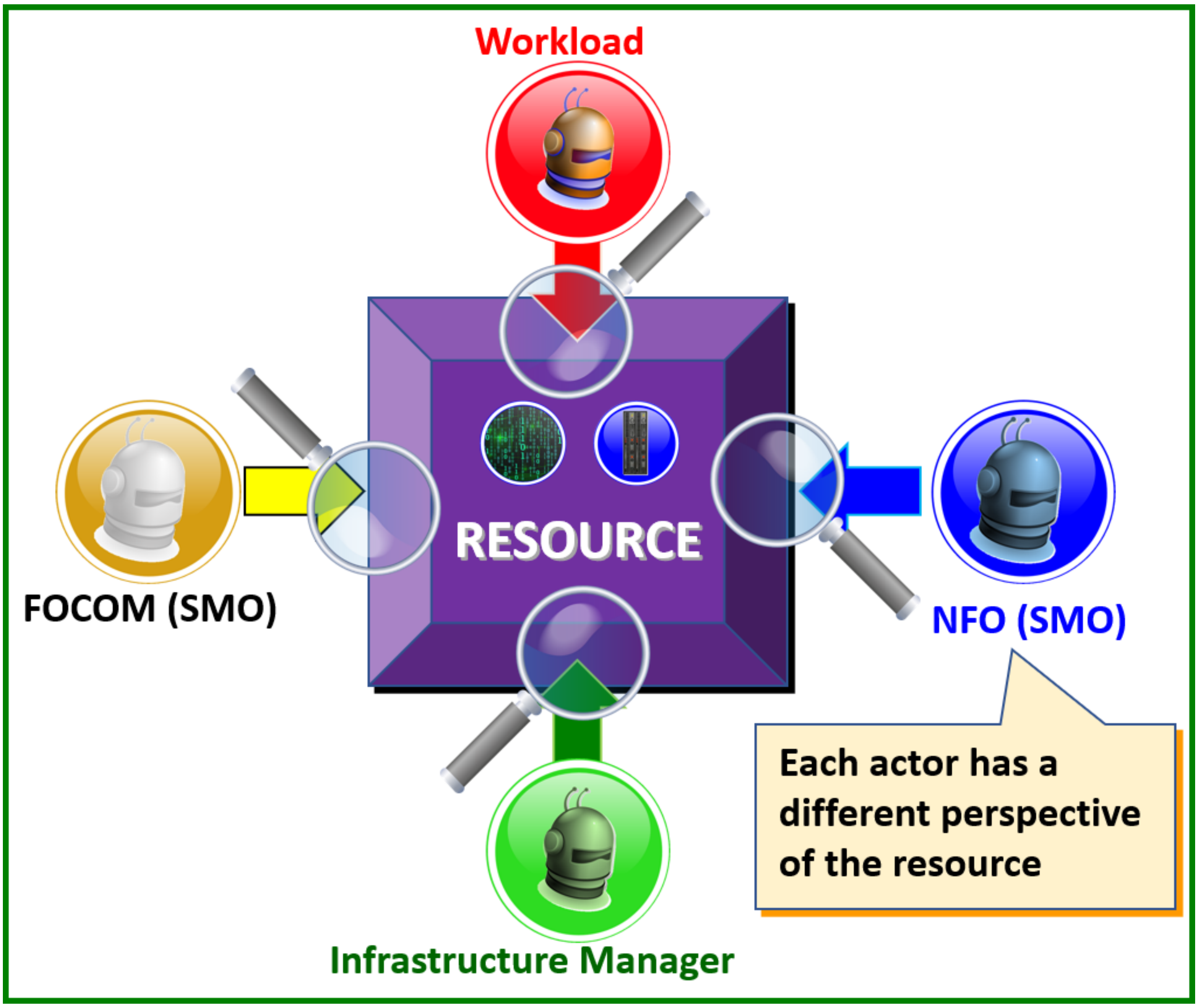


Figure 0-2 Resource Views seen as different actor's perspective

3.4.3.3.1 Cloud Infrastructure Resource View (Green)

The Cloud Infrastructure Resource view is a resource view of a resource that is managed by Cloud internal physical, software and tool components. Cloud Infrastructure Resource can also be composed of other Cloud Infrastructure Resources and services. These can be used to compose or expose Abstracted Resources. The Cloud Infrastructure Resources are typically provided by the Internal, External or Public Cloud Providers.

NOTE: The cloud infrastructure manager (green) is often implemented by a different organization than the FOCOM organization (yellow).

For example, a resource view of the Cloud Infrastructure Resource might have a computer, storage, acceleration, network hardware, and software implementing certain services from a public cloud provider. This is the hardware foundation in any sort of data center that serves as the hardware infrastructure.

3.4.3.3.2 Abstracted Resource View (Yellow)

An Abstracted Resource View is a Resource View of a composed Cloud Infrastructure Resource (with a provisioning reference tag) that is offered through Provisioning. Abstracted Resources thereby map underlying cloud infrastructure resources (act of composing). The Abstracted Resource View is a perspective on the actual Abstracted Resources. This abstraction facilitates assigned users to perform their authorized management and LCM operations for these resources. The SMO sends the objectives to the IMS, which are the SMO requirements for Abstracted Resources that are being requested from IMS. See the figure below on "Resource views: abstracted and assigned resources".

NOTE: The act of composing a Cloud Infrastructure Resource could be as simple as finding suitable resources and tagging it as an Abstracted Resource.

For example, an Abstracted Resource View might contain a set of available computer systems and a K8S cluster that are composed by the cloud infrastructure. The Computer Systems are referenced with an abstraction identifier (assigned by the cloud owner) and the underlying physical compute resources are not exposed. The K8S Cluster is also referenced with an abstraction identifier and the underlying compute, network and storage resources are generally not exposed. Some examples of operations on Abstracted Resources include listings of available resources and booking of available resources. For the logical resource e.g., the K8S Cluster other common operations would also include creation, scaling, update, and deletion of the cluster.

3.4.3.3.3 Assigned Resource View (Blue)

An Assigned Resource View is a Resource View of an Abstracted Resource provisioned to the SMO that can prepare and adapt the resource(s) for Workload placement and consumption including its related FM and PM interfaces. The IMS assigns resources to the cluster and reports to the SMO. The Assigned Resources are the Abstracted Resources that the O-Cloud assigns in order to fulfill a SMO request for resources (see example in Figure 0-3). When requests are fulfilled, the bookkeeping of Assigned Resources and Cloud Infrastructure Resources is done by the O-Cloud IMS. See the figure below on "Resource views: abstracted and assigned resources".

For example, the Assigned Resource View might present one or many K8S cluster(s) and other sorts of resources as an Assigned Resource. Some examples of workload orchestration that can be performed on Assigned Resources are cluster preparation for the workload, placement (deployment) of workloads, and scaling of the workload on the cluster.

For example, consider a rental car as a resource. The Vehicle Identification Number (VIN) is associated with the car. The rental company has a rental agreement number for the same car. The person who rents the car has an entitlement according to the rental agreement using the rental agreement number that is not tied to the VIN. In this case, the rental agreement number is a representation of the car an Abstracted Resource. The renter views the car as an Assigned Resource. At any point in the time, the VIN & agreement number are linked, however, the link can be changed over time. For instance, car #1 breaks down and the rental agreement becomes associated to a new car #2 with a new VIN associated with car #2 all under the same rental agreement number. This illustrates how FOCOM gets Abstracted Resources and hands it to NFO as Assigned Resources.

3.4.3.3.4 Consumed Resource View (Red)

Workload Consumed Resources View is a resource view that describes the workloads software program execution in the CPU; acceleration execution in the dedicated HW accelerator; as well as packet classification, manipulation and forwarding in a packet forwarding device.

For example, the Operating System that consumes CPU instruction resources from the Computer System or Virtual Machine expect to acquire CPU metrics based on information to its OS drivers to offer its built-in performance management services e.g., the common OS TOP commands that indicates how the CPU resources are consumed.

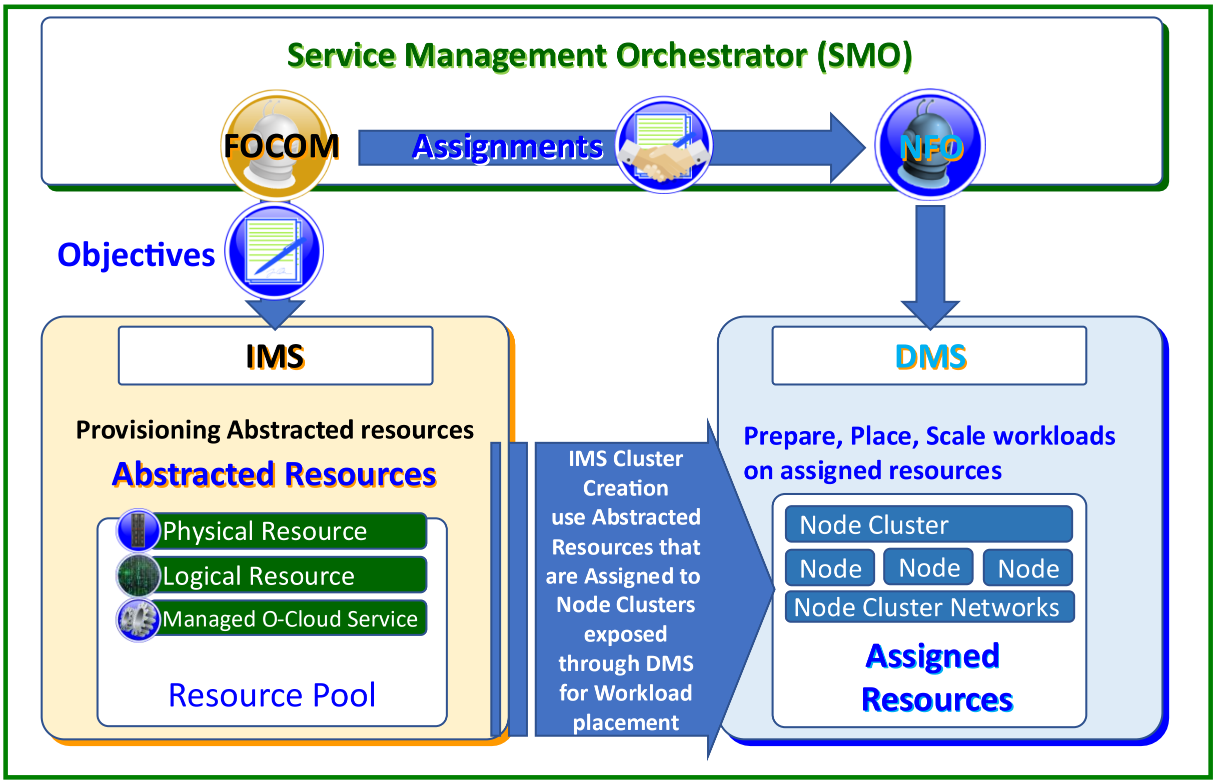

Service Management Orchestrator (SMO)
FOCOM
Assignments
NFO
Objectives
IMS
Provisioning Abstracted resources
Abstracted Resources
Physical Resource
Logical Resource
Managed O-Cloud Service
Resource Pool
IMS Cluster Creation use Abstracted Resources that are Assigned to Node Clusters exposed through DMS for Workload placement
DMS
Prepare, Place, Scale workloads on assigned resources
Node Cluster
Node
Node
Node
Node Cluster Networks
Assigned Resources

Figure 0-3 Resource Views – Abstracted & Assigned Resources Example

3.4.3.4 Hardware Layer Concepts

3.4.3.4.0 Overview

At the foundation of a cloud is the physical hardware with physical computers to execute workloads on and some networks to interconnect the computers with each other. In the real world there are also many different provider organizations that need clear boundaries of their responsibilities e.g., between the cloud infrastructure and the outside world that will be served by the cloud infrastructure and its workloads. There is also a need to house the physical computers, networks, and other related equipment in some physical enclosure.

As introduced in the CADS [i.2] a Cloud Site is a physical place that has Cloud Infrastructure equipment that can be used to provide O-Cloud Resources that are abstracted from the actual physical realization. This enables O-Cloud to lease out O-Cloud Resources through the IMS, based on Cloud Infrastructure physical resources on the Cloud Sites that serves the distributed O-Cloud.

From an SMO point of view, it is important to understand the basic concepts of the hardware layer, not from the realization point of view, but from the O-Cloud Resource locations, capabilities, capacity, availability, and potential disturbances (e.g., at faults or maintenance operations) points of view.

3.4.3.4.1 Computer System

A Computer System is defined to be a physical or composed system capable of performing computations and is also connected to the Site Network Fabric enabling the Computer System to work in a cloud and act as part of a cluster. The Computer System typically includes one or more main CPUs with their memory subsystems, Network Interfaces Controllers (NIC) and sometimes also Storage drives and HW Accelerators. See diagram 4: “Hardware Layer Concepts”. If the Computer System realizes an O-Cloud Node that performs the role of Telecom Slave Clock (T-TSC) LLS-C3 or Telecom Boundary Clock (T-BC) LLS-C1/C2, such Computer System shall embed high speed and low latency NIC(s) with support for PTP Hardware Clock (PHC).

A Computer System can run any major Operating System with or without Virtualization and/or Container support functionality.

For example, a server in a data centre connected to a Site Network Fabric.

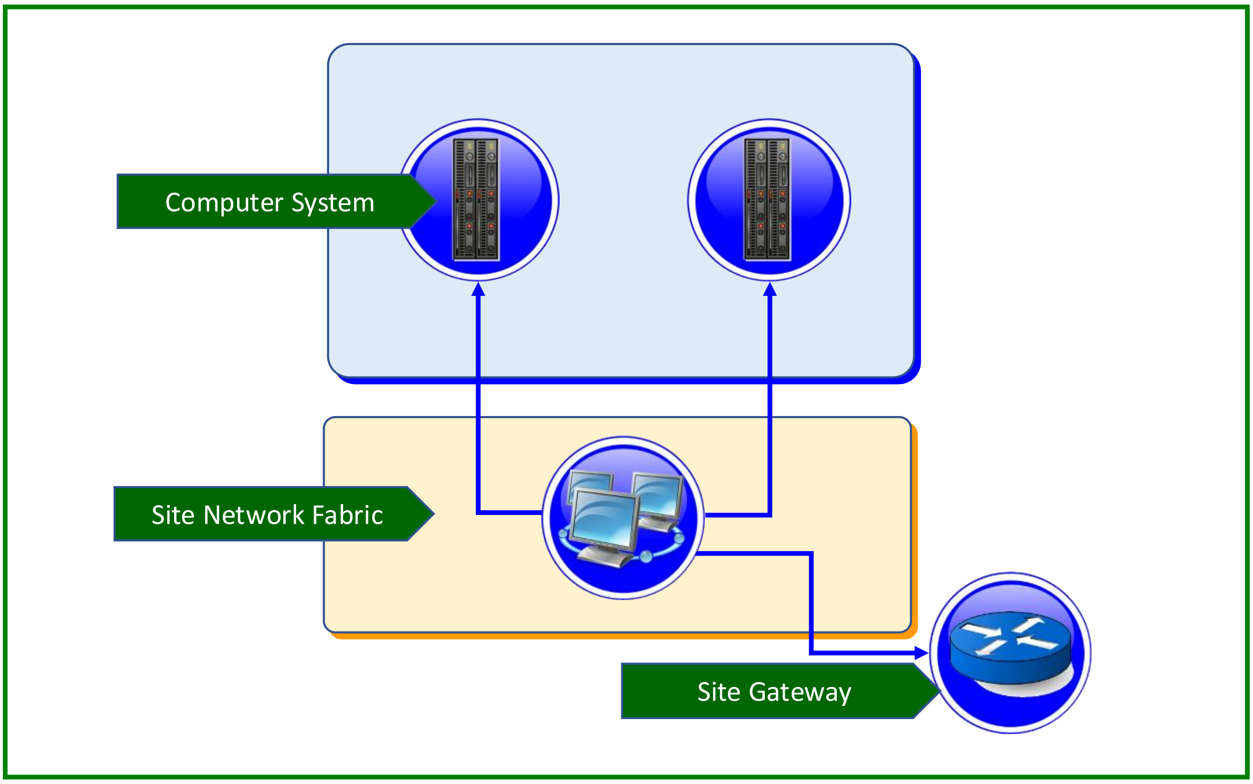


Figure 0-4 Hardware Layer Concepts

3.4.3.4.2 Site Network Fabric

A Site Network Fabric is a physically connected network on a Cloud Site enabling Computer Systems to communicate with each other and with the Gateway(s) connected to networks outside of the Cloud Site.

The Site Network Fabric generally limits how far Layer 2 connectivity can span without being routed or bridged over external routed networks that will normally change the connection characteristics in the form of longer access times. For Full Timing Support networks (refer to clauses 11.2.3 and 11.2.4.2.1 in O-RAN.WG4.TS.CUS.0-R004-v17.00 [39], the Site Network Fabric shall support PTP in order to act as T-BC or Telecom Transparent Clock (T-TC).

For very small Cloud Sites the Site Network Fabric realization can be collapsed to the direct links between the Computer Systems and the Gateways which then are to be treated as fixed connections.

The Site Network Fabric can provision O-Cloud Site Networks as a special type of O-Cloud Resources that can interconnect pooled O-Cloud Resources with each other on the physical level e.g., Computer Systems and Gateways.

3.4.3.4.3 Site Gateway

A Site Gateway (as can be seen in figure 3.4-4) are commonly placed between each Cloud Site and the externally provided transport network to ensure a clear separation of cloud internal and cloud external traffic in terms of security, traffic treatment and operational separation of concerns.

The Site Gateways normally implements routing functionality that separates the Layer 2 domains between the Cloud Sites Site Network Fabric and the external transport domains.

Site Gateways can be implemented as part of the Cloud Infrastructure, O-Cloud or provided by the transport organization. Regardless of where the Site Gateways are implemented, they could be deployed with management either from the Cloud Infrastructure, O-Cloud, or transport organizations.

3.4.3.5 Execution Environment Layer Concepts

3.4.3.5.0 Overview

To make the hardware layer consumable as O-Cloud Nodes, O-Cloud Node Clusters, O-Cloud Site Networks and O-Cloud Gateways, there is a need for Operating Systems and configurations on Computer System, Switch Fabric Network and Site Gateway resources. To enable a distributed O-Cloud and to get traffic in and out of the cloud there is also a need for a configuration of the Gateway(s) that makes it an O-Cloud Gateway.

Beside the Operating System on the Compute Nodes, they can also include cluster software and/or configurations that connects a set of O-Cloud Nodes into a O-Cloud Node Cluster and to a O-Cloud Gateway.

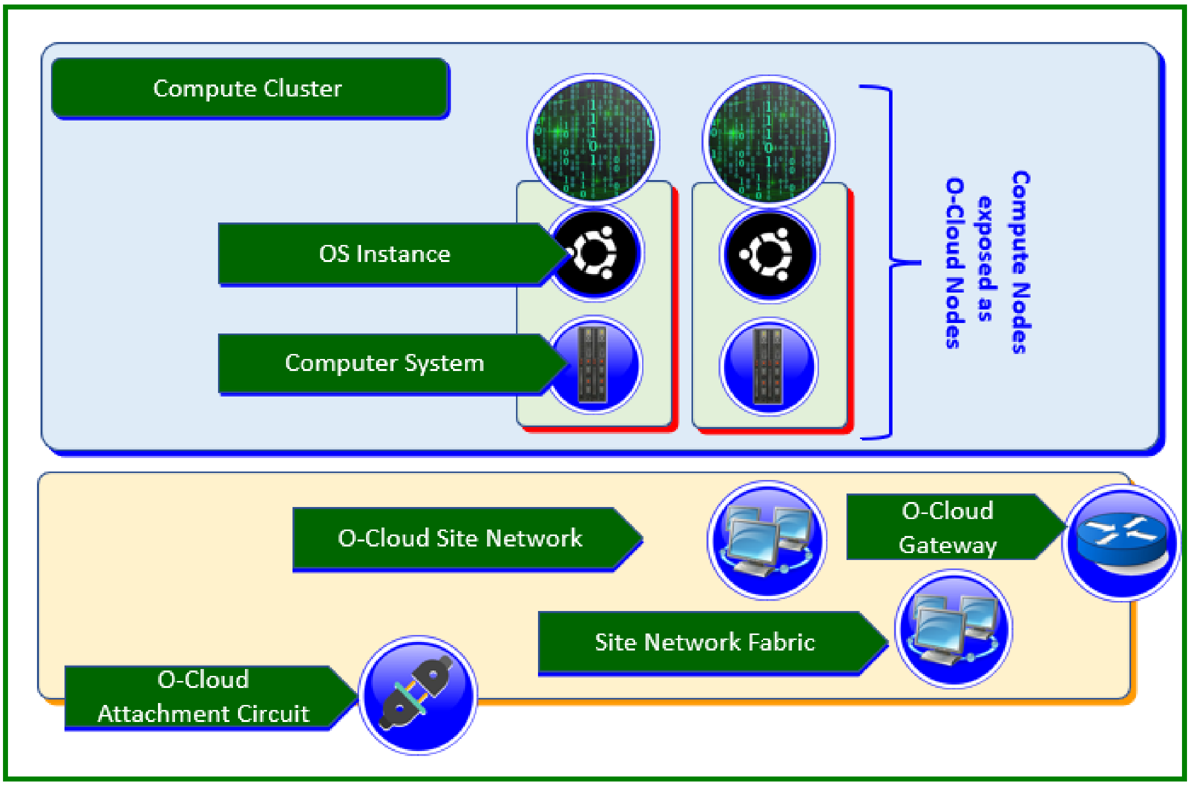


Figure 0-5 Execution Environment Layer Concepts

3.4.3.5.1 Operating System (OS)

An Operating System (OS) is a system software that manages and abstracts the Computer System hardware and software resources as well as provides common services for computer programs such as scheduling and network connectivity. In order to support time synchronization, the OS on a Computer System that performs the role of T-TSC or T-BC shall be configured to synchronize with hardware time stamping by PTP (as an example to relevant OS configuration items, refer to clauses 5.5.1.1.2.1 and 5.5.1.2.2.1 of O-RAN.WG6.CADs **Error! Reference source not found.**).

3.4.3.5.2 Operating System with Virtual Machines

An Operating System with Virtual Machines is an OS that through software or hardware implementation includes the ability to offer multiple Virtual Machines, each acting as a well-separated Computer System.

3.4.3.5.3 Operating System with Containers

An Operating System (OS) with Containers is an OS that include the ability to offer multiple separated name spaces, quotas, and management for Containers.

3.4.3.5.4 Compute Node

A Compute Node is a Computer System with a running managed Operating System that is network connected.

3.4.3.5.5 O-Cloud Node

An O-Cloud Node is an O2ims exposed Abstracted Resource based on the Cloud infrastructure internal Compute Node. (CAD definition: "An O-Cloud Node is a collection of CPUs, Mem, Storage, NICs, Accelerators, BIOSes, BMCs, etc., and can be thought of as a server.").

An O-Cloud Node that participates in the deployment of NF Deployments concerning vO-DU, acting as a master/slave clock, shall be provisioned on a Computer System and OS that has the required PTP support.

3.4.3.5.6 O-Cloud Site Network

An O-Cloud Site Network is a logical network interconnect enabling Compute Nodes to communicate with each other within their cluster and through the O-Cloud Gateway(s) connected to networks outside of the O-Cloud Site including other O-Cloud Sites. The O-Cloud Site Networks can represent a partition of the physically available O-Cloud Site Network Fabrics available networks (sometimes referred to as underlay networks, although it is an ambiguous term since it can refer to any layer that supports another layer). The O-Cloud Site Networks can also be further encapsulated on the O-Cloud Site Networks physical partitions or even on other O-Cloud Site Networks (sometimes referred to as overlay networks, that is neither a distinct term since it can refer to any layer, but the very lowest physical layer).

O-Cloud Site Networks can be provisioned into O-Cloud Node Cluster Site Networks or O-Cloud Node Group Site Networks (often referred to as overlay networks) where they can be used to interconnect the provisioned O-Cloud Nodes inside a O-Cloud Node Cluster on the local O-Cloud Site. The O-Cloud Node Cluster Site Networks can span all the O-Cloud Nodes on the O-Cloud Site while the O-Cloud Node Group Site Network can only span the O-Cloud Nodes within its own O-Cloud Node Group on the O-Cloud Site.

O-Cloud Site Networks, which are used to provision O-Cloud Node Clusters Networks and O-Cloud Node Group Networks in the O-Cloud Site that support NF Deployments with strict time sync requirements, e.g., concerning vO-DU wherein the vO-DU acts as a master/slave clock, shall be provisioned to support T-TC or T-BC.

3.4.3.5.7 Compute Cluster

A Compute Cluster is a selected set of Computer Systems that are physically interconnected through a Site Network Fabric which can be used to setup a O-Cloud Node Cluster (as defined below).

3.4.3.5.8 O-Cloud Node Cluster

An O-Cloud Node Cluster is a set of O-Cloud Site Network interconnected Compute Nodes of a specific type. The following four types of O-Cloud Node Clusters are defined:

1. **Bare Metal Container Cluster –** A Bare Metal Container Cluster is a set of network-connected computer systems with their individual operating system instances that supports containers in a cluster configuration.
2. **VM-based Container Cluster –** A VM-based Container Cluster is a set of network-connected Virtual Machines with their individual guest operating system instance that supports containers in a clustered configuration.
3. **VM Cluster –** A VM Cluster is a set of network-connected Computer Systems with their individual operating instance that supports virtual machines in a cluster configuration.
4. **OS Cluster –** A OS Cluster is a set of network-connected Computer Systems with their individual operating system instance that supports a cluster configuration.

An O-Cloud Node Cluster that participates in the deployment of NF Deployments concerning vO-DU, acting as a master/slave clock, shall be provisioned with O-Cloud Nodes whose Computer System and OS have the required PTP support (see clause 3.4.3.5.5).

3.4.3.5.9 O-Cloud Gateway

An O-Cloud Gateway can also be used to encapsulate, bridge, or stitch the O-Cloud Site Networks in different O-Cloud Sites via transport networks e.g., for realizing Distributed O-Cloud Node Cluster Networks.

3.4.3.5.10 O-Cloud Attachment Circuit

An O-Cloud Attachment Circuit (OCAC) is a logical connection enabling connectivity of Site Networks deployed within the O-Cloud Site to outside of the O-Cloud Site. In the deployment scenario where an O-Cloud Site includes the gateway functionality, one or more OCACs can exist on the O-Cloud Gateway.

3.4.3.5.11 O-Cloud Bearer

An O-Cloud Bearer is a physical or logical link that establishes connectivity between the O-Cloud Site and transport networks or other networks outside the O-Cloud Site. It’s possible to carry one or more OCAC over the same O-Cloud Bearer.

3.4.3.6 Local and Distributed O-Cloud Node Clusters

3.4.3.6.0 Overview

Node Clusters can be deployed local to one Cloud Site or distributed over multiple Cloud Sites, but they are always within one O-Cloud.

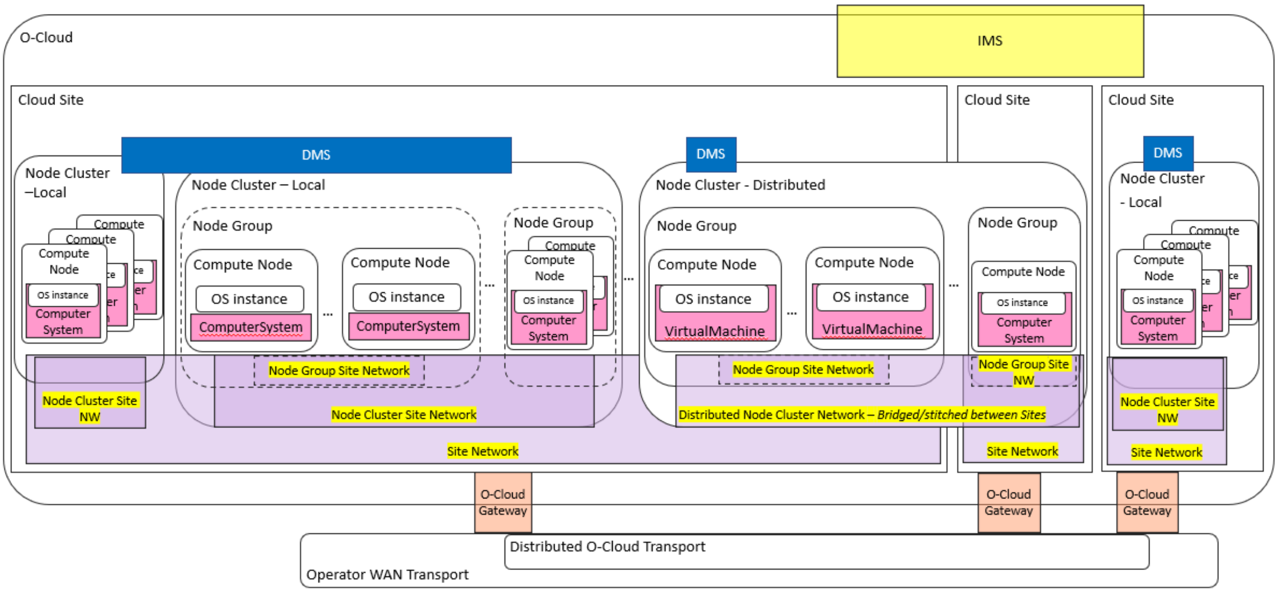


Figure 0-6 O-Cloud example with Local and Distributed Node Clusters

3.4.3.6.1 Distributed O-Cloud Node Cluster Networks

Distributed O-Cloud Node Clusters that span multiple O-Cloud Sites also span an O-Cloud Gateway routing function that can connect the O-Cloud Node Cluster Site Network traffic on the local O-Cloud Site to an external transport network. The O-Cloud Gateway routing functions in each O-Cloud Site where the distributed O-Cloud Node Cluster has O-Cloud Nodes, are configured to stitch, or bridge each O-Cloud Site’s local O-Cloud Node Cluster Site Networks with each other. The O-Cloud Gateway router function can also be configured to stitch or bridge selective O-Cloud Node Group Site Networks in between O-Cloud Sites.

The external transport network could be a dedicated Distributed O-Cloud Transport or a configured partition of the Operators WAN Transport network.

For deployments of distributed O-Cloud Node Clusters that will depend on the local traffic characteristics of the local O-Cloud Site the O-Cloud Nodes on each O-Cloud Site ought to be provisioned into one or more O-Cloud Node Groups if the clusters have a scheduling mechanism e.g., K8s. O-Cloud Node Group Site Networks can be used where specific O-Cloud Site local traffic treatment or isolation is desirable.

3.4.3.7 Initiation Configuration Set Concept

Several Cloud Infrastructure and O-Cloud Resources have to be initiated or initially configured before they can be used for anything. This initialization could be done at installation, power-on, boot-up, start-up or as a re-configuration when the resource is not in-service.

Many of these initial configurations have dependencies of each other and might be destructive to its functionality, characteristics or even expected lifespan, e.g., an incorrect combination of configurations could permanently destroy resources. Examples of such configuration parameters are found in servers BIOS settings and networking devices port configurations or setup of physical layer parameters.

Some initial parameters only take effect during start-up and others could later be changed during run-time. Examples of parameters generally used only at start-up are specific hardware mode configurations and links to firmware and software images supplied by the Cloud Infrastructure or O-Cloud. Examples of run-time changeable configurations are port speeds of IO modules. An important parameter for a Compute Resource to be useful is the OS image (possibly including Cluster SW and other functions) that will be booted for an O-Cloud user. The image enables the O-Cloud Resource to be used and exposed as a O-Cloud Node in a O-Cloud Node Cluster. Other important initiation configuration concerns the settings on the hardware (such as Computer Systems and Site NW Fabrics) and OS layers to support time synchronization.

To allow the O-Cloud users a safe yet flexible way to setup these initial configuration parameters in a coherent way, the O-Cloud has a concept of an Initiation Configuration Set (ICS) artifact that should be a consistent, validated, static and safe set of configuration parameter available in the O-Cloud. The ICS artifact content can be applied to the O-Cloud Resources at each start-up occasion. ICS artifacts are likely rather specific for each ResourceType and should have an association that indicate what ResourceType or ResourceTypes it is validate for.

The resulting started O-Cloud Node keeps a reference to the used ICS that can be used to understand the static initiation parameters that can only be set at start-up. Configuration parameters initialized by the ICS, that can be modified in run-time towards the O-Cloud Node, will not change in the ICS since that is a static artifact. These parameters in the ICS could consequently not be used as reference to what is currently running in the O-Cloud Node.

When the ICS is applied to the Resource at initialization it will result in a new set of capabilities. The user of the O-Cloud Resources will have to keep within the capabilities (hard limitations) of the installed and available O-Cloud Resources but have the opportunity to apply different ICS to tailor these Resources at each start-up to take into account the new set of capabilities and capacities.

Each ICS are linked to the ResourceType it can be applied to, but there can be multiple ICS that support the same ResourceType enabling the O-Cloud user a selection of soft limitations to be applied at start-up. If the O-Cloud supports, the use of ICS then O2ims need to expose available and referenced ICS.

Where a ResourceType supports multiple separated or hierarchical restart or start-up domains the ResourceType could select to divide up its total set of start-up configuration parameters in separate ICS that can be individually applied for each such domain. This pattern enables a ResourceType to perform a partial reconfiguration of the resource during run-time, but it can also be used for other practical reasons. For fixed pre-configured Resources there might not even be an ICS available in the O-Cloud.

3.4.3.8 Managed Infrastructure Template Concept

The Managed Infrastructure Template (MIT) concept enables repetitive creation and updates of managed O-Cloud Resources. Examples of such managed O-Cloud Resources are O-Cloud Node Clusters and Infrastructure Resources that are provisioned for the O-Cloud client (SMO) and managed by the O-Cloud.

The MIT contains a parameter schema enabling an O-Cloud client to determine the mandatory and optional parameters to be supplied with a client’s provisioning request for a set of managed O-Cloud Resources. The MIT can also contain a set of known target characteristics that enable the O-Cloud client to understand what capabilities and capacities the resulting set of managed O-Cloud Resources will have when they are provisioned. These characteristics are intended to support the O-Cloud client in their selection of what MIT to choose for a specific provisioning request.

The O-Cloud has the MIT source-of-truth, and a set of O-Cloud internal MIT related content used to control the realization of managed O-Cloud Resources when provisioning requests are received. The MIT with a valid set of parameters serves as the declarative target configuration of a provisioning request.

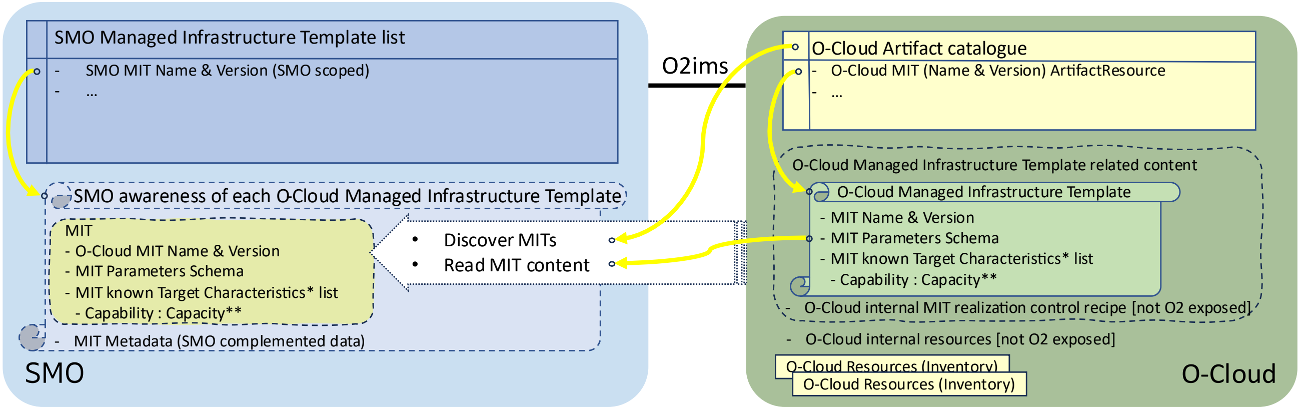
Templates are a common concept used for declarative target of clusters in various initiatives, e.g., in clause 6.2 of ETSI GS NFV-IFA 036 [19] for the case of container infrastructure clusters.

SMO selects a MIT representing a desired set of characteristics of managed O-Cloud Resources, selects values for that MIT's defined parameters and sends a provisioning request to the O-Cloud. The known target characteristics and a parameter schema are part of the MIT. The O-Cloud exposes the MITs to the SMO as MIT Artifact Resource objects on O2ims hence enabling the SMO to discover them and query them about their attributes. The parameter schema in the MIT defines the parameters, their types, their allowed value ranges and whether they are mandatory or optional to supply with a provisioning request that references this MIT.

When the SMO issues a provisioning request referencing a MIT in the O-Cloud and, where applicable, includes a valid set of parameter values, it triggers the O-Cloud to start convergence towards the MIT declarative target of the requested managed O-Cloud Resources. The O-Cloud applies its O-Cloud internal MIT realization control recipe on O-Cloud Inventory Resources and O-Cloud internal resources that enable the O-Cloud to converge to the declarative target configuration of the managed O-Cloud Resources.

The O-Cloud internal MIT realization control recipe provides the O-Cloud with required information to build the managed O-Cloud Resources, including configuration, topology and structure of O-Cloud Inventory Resources and/or O-Cloud internal hardware and software resources. The MIT can also include the managed O-Cloud Resources known characteristics to be made available in their respective resulting target managed O-Cloud Resource objects. A standardized set of managed O-Cloud Resource capabilities used for the Node Cluster creation and updates are defined in Annex D. The resulting managed O-Cloud Resources will be built and reconciled in the O-Cloud, whereafter it is configured e.g. as an O-Cloud Node Cluster and/or a set of O-Cloud Infrastructure Resources reflecting the selected MIT with its known characteristics. When a provisioning request is fulfilled and the managed O-Cloud Resource has converged to its declarative target, they will have their own set of detailed characteristics exposed through their respective resource objects e.g. Node Cluster Resources and Infrastructure Resources.

An O-Cloud Node Cluster instantiated based on a MIT will have specific capabilities and capacities that can make it more suitable or less suitable for orchestrating certain NF Deployments on that O-Cloud Node Cluster. Supported capabilities can be expressed by labelling.


SMO Managed Infrastructure Template list
- SMO MIT Name & Version (SMO scoped)
- ...
SMO awareness of each O-Cloud Managed Infrastructure Template
MIT
- O-Cloud MIT Name & Version
- MIT Parameters Schema
- MIT known Target Characteristics* list
- Capability : Capacity**
- MIT Metadata (SMO complemented data)
Discover MITs
Read MIT content
SMO
O2ims
O-Cloud Artifact catalogue
- O-Cloud MIT (Name & Version) ArtifactResource
- ...
O-Cloud Managed Infrastructure Template related content
O-Cloud Managed Infrastructure Template
- MIT Name & Version
- MIT Parameters Schema
- MIT known Target Characteristics* list
- Capability : Capacity**
- O-Cloud internal MIT realization control recipe [not O2 exposed]
- O-Cloud internal resources [not O2 exposed]
O-Cloud Resources (Inventory)
O-Cloud Resources (Inventory)
O-Cloud

*) Target Characteristics not depending on parameters

**) Capabilities and Capacities may be O-RAN well-known (standardized), O-Cloud Vendor defined, or NW Operator defined through MIT

Figure 0-7 Managed Infrastructure Template related parts and references in SMO and O-Clouds

The intended usage of the MITs is that a provisioning request with a MIT reference together with its populated parameters represents a target configuration of a set of managed O-Cloud Resources. That target configuration has a set of known characteristics that are well suited for the deployment of the intended target workloads (e.g., NF Deployments) through the fulfilment of their required capabilities, such as NF Deployments realizing O-RAN Cloudified NF that need support for time synchronization.

An example of using a MIT for Cluster provisioning and configuration settings to an O-Cloud Node Cluster are functions related to time synchronization. In such an example, the provisioning request, with referenced MIT and needed parameterization defines the requirements for O-Cloud Site NWs, and/or O-Cloud Gateways to support PTP transmission, and O-Cloud Node Cluster to include O-Cloud Nodes whose OS is set to handle hardware time stamping for PTP.

For the creation of a MIT and its O-Cloud related content, the following aspects need to be considered: O-Cloud exposed and internal resources, their internal connectivity, software, and configuration parameters as well as the requirements of the workloads that might be deployed on the resulting set of managed O-Cloud Resources. The MIT is immutable and can only be created or deleted. It is referenced by a Name and a Version that together shall be unique within an O-Cloud. The MIT Name is intended to enable the Network Operator to keep track of what main purpose the MIT has, and the MIT Version is a way to allow the Network Operator to have several versions of the MIT with a specific Name. When a MIT and its related content is created for an O-Cloud and is ready for usage in the O-Cloud, it is exposed as an O-Cloud Artifact Resource. This MIT

Artifact Resource provides the parameters schema that SMO needs to inject suitable parameters, and a representation of the resulting known managed O-Cloud Resources target characteristics, such as, but not limited to their capabilities and capacities. SMO can detect any new MIT over the O2ims through polling or subscribed change notifications of the MIT Artifact catalogue and can then read each MIT when it is required. The SMO can build up its own representation of each O-Cloud's MIT and complement these with relevant metadata such as where in the O-Cloud the MIT is supported. This information enables the SMO to understand the applicability of a MIT and where it can be used to deploy its targeted set of managed O-Cloud Resources.

SMO keeps track of MITs to enable its orchestrators and other functions to understand what MITs are available in each O-Cloud, and what known characteristics could be realized by a provisioning request referencing a specific MIT and other relevant metadata about the MITs. This information is required to enable SMO to select a relevant MIT for the set of targeted managed O-Cloud Resources that can serve the intended workloads.

When a SMO requests a new or updated set of managed O-Cloud Resources it references the desired MIT in the O2ims ProvisioningRequest along with a complete set of valid parameters for the selected MIT. The O-Cloud will then reconcile the set of managed O-Cloud Resources to the declarative target given by the referenced MIT and SMO supplied parameters. The managed O-Cloud Resources and its constituent O-Cloud Resources and topologies are reflected through their respective objects using the O2ims Inventory service that SMO can poll or get notifications about.

NOTE 1: The method of how the MIT is inserted in the O-Cloud is not described in the present document.

NOTE 2: The method of how the MIT is created and deleted is not described in the present document.

NOTE 3: The method of how the MIT is created and tested is not described in the present document.

NOTE 4: The SMO represented MIT characteristics such as Capabilities and Capacities are not described in the present document.

3.5 O-Cloud Software Management

Software Management should be a priority as without proper management unnecessary risks may be taken. Software Management ensures security, cost management and software support. There are many benefits to Software Management, of which the main benefits are:

- Prevents unauthorized software from being installed
- Maintains a catalog of authorized software and its versions
- Provides visibility into what software and version is being used
- Provides a better view of which software products and versions are vendor supported

In O-RAN, from an O-Cloud perspective, there are four types of software which needs to be managed on the O2 Interface:

- The IMS software within an O-Cloud
- The DMS software within an O-Cloud
- The O-Cloud Platform software
- The O-RAN Cloudified NF software is the software implementation of O-RAN NFs which is capable of running over the O-Cloud

3.5.1 General Capabilities

Pre-requisite:

- O-RAN Cloudified Network Function Software Package is onboarded to SMO with required information, the details are not documented in the present document
- From an O-Cloud perspective the O-RAN Cloudified Network Function Software Image includes the underlying software executable image, image properties/metadata such as version

- IMS Software Version image(s) is available in or from SMO with required information, the details are not documented in the present document
- DMS Software Version image(s) is available in or from SMO with required information, the details are not documented in the present document
- Security/Certificates, the details are not documented in the present document

Business Capabilities:

- Network Operations shall be able to perform the Software Image Management of O-Cloud Platform, IMS, DMS and O-RAN Cloudified Network Function

Functional Capabilities:

- O-Cloud shall provide Add Software Images of O-Cloud Platform, IMS, DMS and O-RAN Cloudified Network Function to O-Cloud repository
- O-Cloud shall provide Delete Software Images of O-Cloud Platform, IMS, DMS and O-RAN Cloudified Network Function from O-Cloud repository
- O-Cloud shall provide Update Software Images of O-Cloud Platform, IMS, DMS and O-RAN Cloudified Network Function to O-Cloud repository
- O-Cloud shall provide Query Software Images of O-Cloud Platform, IMS, DMS and O-RAN Cloudified Network Function from O-Cloud repository
- O-Cloud shall provide Software Image properties information of O-Cloud Platform, IMS, DMS and O-RAN Cloudified Network Function.
    - SoftwareImageId, Vendor, and Version
- IMS can obtain the requested IMS Software, perform needed pre-check(s), and perform the update
- IMS can obtain the requested DMS Software, perform needed pre-check(s), and perform the update

3.6 O-Cloud Life Cycle Management

O-Clouds are a set of hardware and software components that provide cloud computing capabilities to execute O-RAN Cloudified Network Functions. These O-Clouds can be deployed centrally or at edge based on the deployment scenario and this typically involves cost and complexity. The goal of O-Cloud Life Cycle Management is to reduce this cost and complexity by orchestrating the deployment and management of the O-Clouds.

In O-RAN the O-Cloud Life Cycle Management will provide the following capabilities:

- Deploy
- Registration
- Scale

The objective of Deployment is to provide automated provisioning of the O-Cloud Infrastructure.

The objective of Registration is to register an O-Cloud and make it available for deployments. The SMO provides information for the registration of O-Clouds, this includes security handshake required for the trusted connection, the O-Cloud IMS endpoint, and the O-Cloud ID which is used to correlate the SMO inventory record with the deployed cloud instance.

After the registration of an O-Cloud with the SMO, the following information but not limited to can be discovered:

- O-Cloud (IMS) provided Inventory services
    - List of All Resource Pools in the O-Cloud
    - Attributes of a specific O-Cloud
    - List of all resources of an O-Cloud Pool

  - Attributes of each O-Cloud Resource
  - List of all DMS
- O-Cloud (DMS) provided Inventory services
  - List of Locations Supported
  - For a given location the Capabilities supported, such as:
    - Descriptor types
    - Technology types
    - Accelerator types
  - For a given location the Capacity of the location
  - For a given location the Availability of the location

The Scaling is not present in the current document.

3.7 O-Cloud Deployment Life Cycle Management

3.7.1 Overview

RAN networks are in an evolutionary journey towards NFV/SDN, Openness and Network Function disaggregation. This journey enables the CSP's quest for future network and operational efficiency, where they can rapidly deliver new services and update the existing services with less cost than today. The goal is to reduce the time to market for a new service launch.

To achieve this end state of higher service agility and lower operational economics, the CSPs are unambiguous in their support for a new network operational framework. The life cycle management of RAN network functions is an important consideration for this new network operational approach.

In O-RAN, the O-Cloud will provide Deployment Life Cycle Management of NF Deployments. The following capabilities will be provided:

- Instantiate, which deploys NF Deployments on O-Cloud with necessary O-Cloud resources.
- Terminate, which terminates NF Deployments on O-Cloud with releasing associated O-Cloud resources.
- Scale, which scales functional behavior and resources of NF Deployments to support services. There is different variance in scaling:
  - Horizontal Scaling provides dynamic functional scaling of NF Deployments' behaviour by Scale In/Out. This scaling can have different invoke actors such as O-Cloud and SMO.
  - Vertical Scaling provides dynamic resource scaling of NF Deployments by Scale Up/Down.
- Heal, which provides the ability to recover/mitigate the NF Deployments' abnormal behaviour in the network.

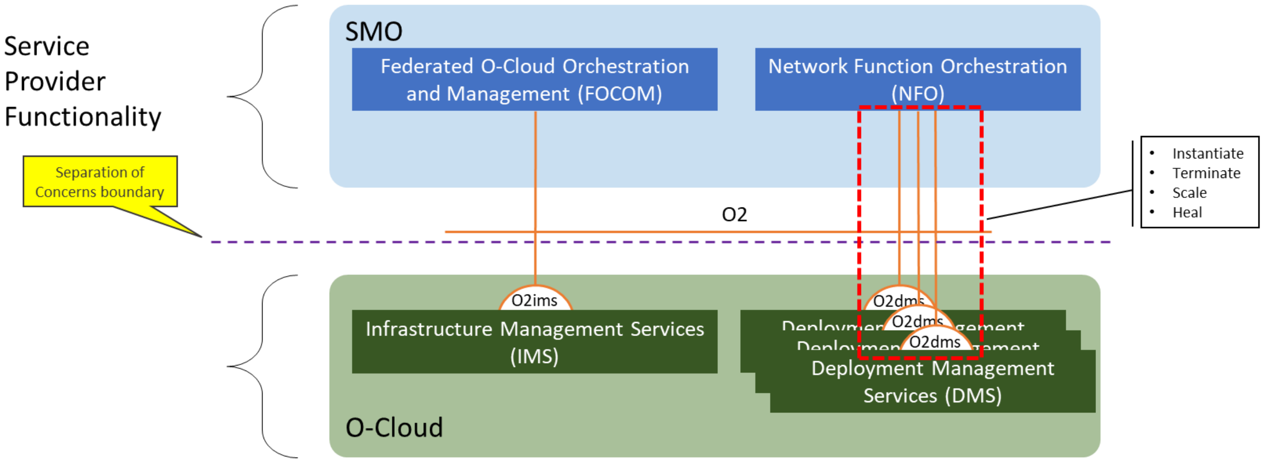

Service Provider Functionality
Separation of Concerns boundary
SMO
Federated O-Cloud Orchestration and Management (FOCOM)
Network Function Orchestration (NFO)
Instantiate
Terminate
Scale
Heal
O2
O2ims
O2dms
O2dms
O2dms
Infrastructure Management Services (IMS)
Deployment Management Services (DMS)
O-Cloud

Figure 0-1 O2 Deployment LCM Services

3.7.2 General Capabilities

Business Capabilities:

- The SMO provides the capability for Network Operations to initiate the lifecycle management functional capabilities

Functional Capabilities:

1. O-Cloud supports Instantiating an NF Deployment instance
2. O-Cloud supports Terminating an NF Deployment instance
3. O-Cloud supports Horizontal Scaling (in and out) of an NF Deployment instance
4. O-Cloud supports Healing of an NF Deployment instance
5. O-Cloud supports Querying information about an NF Deployment instance
6. O-Cloud supports Querying status of LCM operations
7. O-Cloud supports Upgrading of any or all components of an NF Deployment instance

3.7.3 Deployment Management Services

The O-Cloud provides services over the O2dms interface to allow an SMO to trigger the functional capabilities specified in Section 0. As allowed by the O-RAN Cloud Architecture and Deployment Scenarios [12] specification, O-Cloud implementations can incorporate a multitude of virtualization technologies (e.g., VMs, containers, bare metal) and technology stacks which have major differences from each other. To account for these differences, each O-Cloud might have its own specific O2dms service API profile(s). An O2 DMS specification for a specific profile describes the API provided by the profile and how to use them to achieve the capabilities listed in Section 0. An SMO is intended to be able know which profile it is using and orchestrate according to description provided in the O2dms service API profile(s). The profile specific APIs will be defined in each stage 3 O2dms specification. The SMO needs to understand which O-Cloud capabilities are available in the O-Cloud, and what capabilities are needed by NF Deployments, that realize, all or part of, a Cloudified NF, so it can match the infrastructure requirements for the NF Deployments with the capabilities available in the different O-Cloud Node Clusters. A standard set of O-Cloud capabilities used for the deployment purposes of the NF Deployments are defined in Annex D.

The O2dms specification will contain the following information:

a) O-Cloud profile scope and technology stack specifies what types of technologies (e.g., VMs, containers) the profile applies to and specifies particular software systems (e.g., K8S, Prometheus, etc.) and versions that are part of the stack.

b) O2dms functional capabilities description specifies how the APIs provided by the O-Cloud Profile are used (e.g., interaction flow) to implement the functional capabilities specified in Section 0. To achieve these capabilities, dependent on the specific profile the SMO can do one or more of the following:
   1. Initiate the procedure via a single DMS API profile call.

2. Orchestrate using several DMS API profile calls.
3. Relegate responsibility to the O-Cloud by leveraging O-Cloud capabilities, SMO does nothing to manage these capabilities.

c) O-Cloud profile API specification: specifies what APIs are provided by an O-Cloud conforming to the O-Cloud Profile.

d) O-Cloud profile data model specifies the actual data elements used by the API.

The O-Cloud Profiles form the basis for the protocol and data model O2dms interoperability – any O-Cloud implementation which supports a particular O-Cloud Profile is expected to interoperate with any SMO implementation that also supports the same O-Cloud Profile but not with an SMO that only supports a different O-Cloud Profile.

3.8 O-Cloud Fault Management Concepts

3.8.0 Overview

This section details general concepts related to IMS & DMS fault management. Fault management is a basic part of OA&M functions. The concepts described here will motivate the development of the O-Cloud O2 IMS & DMS Fault services and information modeling.

The O-Cloud may be a confederation of resource pools comprised of various compute, storage, and networking resources. These might be provided by different public cloud providers. The public cloud provider is defined as computing services offered by third-party providers over the public Internet, making them available to anyone who wants to use or purchase them. Thus, it is important to outline some basic principles that can apply and are relevant to disparate vendors and providers.

The following basic concepts are considered for IMS & DMS fault management:

3.8.1 Fault Information Model

A fault information model is expected to be developed in the O2ims information model. As a general concept, an information model provides a common lexicon, terminology, and model to serve as a basis for coordination between a group of actors. It facilitates communications between groups and aligns their contributions. The fault information model is applicable for the resources & resource pools as defined in the infrastructure information model.

3.8.2 Subscription

It is expected that there will be a subscription mechanism between a client and an agent. The client would pass an endpoint where notifications should be sent to. It is expected that multiple clients can be supported. Procedures are expected to be defined in the O2 specifications for IMS and DMS. Subscriptions originate from the client (e.g., SMO).

For example, the concept of a subscription allows a client, such as the SMO, to register for call-backs to the O-Cloud to receive Alarm Notifications.

3.8.3 Notifications

Alarm Notifications originate from the IMS within the O-Cloud. There is a relationship between the Subscription and the Notification. When a condition occurs on an O-Cloud Resource which causes the current alarm list to change, this triggers an evaluation of the Alarm Subscription criteria by IMS. If deemed relevant by the Subscription criteria, an event is issued towards clients that have subscribed to receive such a type of Alarm Notification.

For example, a user could receive alerts on an Alarm dashboard GUI based on O-Cloud IMS Alarm Notifications.

3.8.4 Parallel Reporting

A fault occurrence on an O-Cloud Resource may be reported by IMS to consumer(s) as an Alarm Notification. DMS resource faults are associated with a workload. It is possible that multiple resources can associate with a workload, whereby one fault might trigger an Alarm Notification on IMS, DMS and O1. Due to the nature of the subscribe/notify mechanism for Alarm reporting, it is possible to: (1) have multiple publishers sending to a single consumer and (2) conversely a single publisher that can send to multiple consumers. The interfaces for subscription & notification of Alarm reporting are: O2ims, O2dms, O1.

For example, it is conceivable that a vDU application might report a fault on O1, and a related physical server allocated to a workload of a vDU would raise a DMS fault which might also have an infrastructure Alarm raised on IMS.

3.8.5 Simple Devices

Collectively, in the O-Cloud it is conceivable that many devices are simple devices that do not have a full suite of OAM S/W.

For example, suppose a network interface card (NIC) is a resource that is a simple device, and if that NIC encounters a fault it might not be able to register, store, or log faults. It is expected that O-Cloud infrastructure could manage these kinds of simple devices to be able to fulfill some of the other basic fault concepts described in this section. The O-Cloud infrastructure works with the IMS to report these faults.

3.8.6 Source of Truth

In an O-Cloud IMS context, it is expected that the "*source of truth*" will be the resources of the O-Cloud itself. When an Alarm has been reported to a client, such as the SMO, the fault data is replicated within the Alarm Notification to provide details and information of the associated fault. Thus, the fault information now exists in multiple places both in the IMS and the entity receiving the Alarm Notification. Which element then has the "true" information should the network or elements be disconnected or faulty? After connectivity returns, how would the SMO reconcile its view of the existing Alarms against the current Alarms on the O-Cloud? The answer is for IMS to query the *source of truth* (O-Cloud Resource) for the most current faults.

For example, if a resource in the O-Cloud goes out of service, and IMS had previously sent an Alarm Notification, and then later the O-Cloud Resource returns to service, potentially the client (SMO) and the O-Cloud will have different Alarm information. But the O-Cloud Resource is the source of truth, and it is expected that client (IMS) would query the O-Cloud Resource for the current information and reconcile that with client data and if deemed necessary based on the subscription criteria, an event is issued towards clients that have subscribed to receive such a type of Alarm Notification.

3.8.7 Alarm Dictionary

An Alarm Dictionary is a reference which details the alarms that can be emitted from that O-Cloud for one or more O-Cloud Resource(s). The set of Alarm Dictionaries define the failure/fault conditions that would be reported by the IMS in the O-Cloud so that the SMO could act on those. The IMS delivers the set of Alarm Dictionaries to the SMO, for example during the process of 1) O-Cloud registration, 2) IMS Software update 3) O-Cloud Resource addition and 4) O-Cloud Resource software update. The desirable goal is commonality of the Alarm Dictionaries across O-Clouds even though they will differ to some degree by implementation.

For example, the Alarm Dictionary could be used for a variety of purposes. It might help a user decode the meaning of an Alarm Notification. As another example, an Alarm Dictionary could allow for fault signatures or patterns to be identified which may constitute a condition which requires a coordinated response.

3.8.8 Queries

In the IMS case, the SMO (client) can issue an Alarm Query towards to IMS (publisher) related to the O-Cloud. In the SMO to DMS case, the SMO (client) can issue a specific query related to xApp/NF deployments (e.g., workloads) from the DMS (publisher). In the IMS case, an Alarm Query is a message which is expecting a response message from the IMS which matches the selection criteria. In the DMS case, the Alarm Query is expecting a response from the DMS.

For example, the client may request status of a specific type of Alarm, or new active alarms conditions. The O-Cloud would reply to the query by responding with data based on the parameters of the query (e.g., IDs or severity).

3.8.9 Filtering

It is expected that the client, such as the SMO, can query for specific kinds of Alarms or groups of Alarms through a filter. In addition, filtering can also be applied to Subscriptions to determine if Notifications should be set.

For example, it might request from the O-Cloud all the major severity alarms from a particular O-Cloud. The O-Cloud would select records based on the query pattern and send the details according to the Notifications. This concept will guide the IMS information model development of procedures that can support filtering.

3.8.10 Void

Void.

NOTE: The contents of this clause (Fault Logging) has moved to clause 3.8.19.

3.8.11 Alarm Synchronization

There are many potential situations where a client and source of truth might become disconnected or disrupted. In this case, the client will need to synchronize with the source of truth after it is able to reconnect.

For example, network fault might prevent communication between the SMO (client) and IMS resources in the O-Cloud. By the time connectivity has been restored, many new alarm conditions might have arisen and logged. This would motivate the SMO to resynchronize to get the current alarms again. A variety of mechanisms might be used to trigger alarm resynchronization. It might also trigger an upload of the most recent fault logs.

3.8.12 Alarm Correlation

As described in Parallel Reporting, it might be possible that a fault condition might result in a variety of different conditions or be reported in any combination of IMS, DMS and O1 Alarm Notifications. It is not the job of the O-Cloud IMS fault management to correlate faults between the application/NF and O-Cloud Resources. Thus, it would be up to higher layers of fault management to correlate disparate Alarm events to identify common root causes.

For example, the SMO may receive an Alarm Notification for a fault with a root cause of a network issue among the O-Cloud Resources which manifests itself in a deployment and application fault as well. Thus, the SMO might receive three notifications over O2ims, O2dms and O1. It might correlate these Alarms to a common root cause.

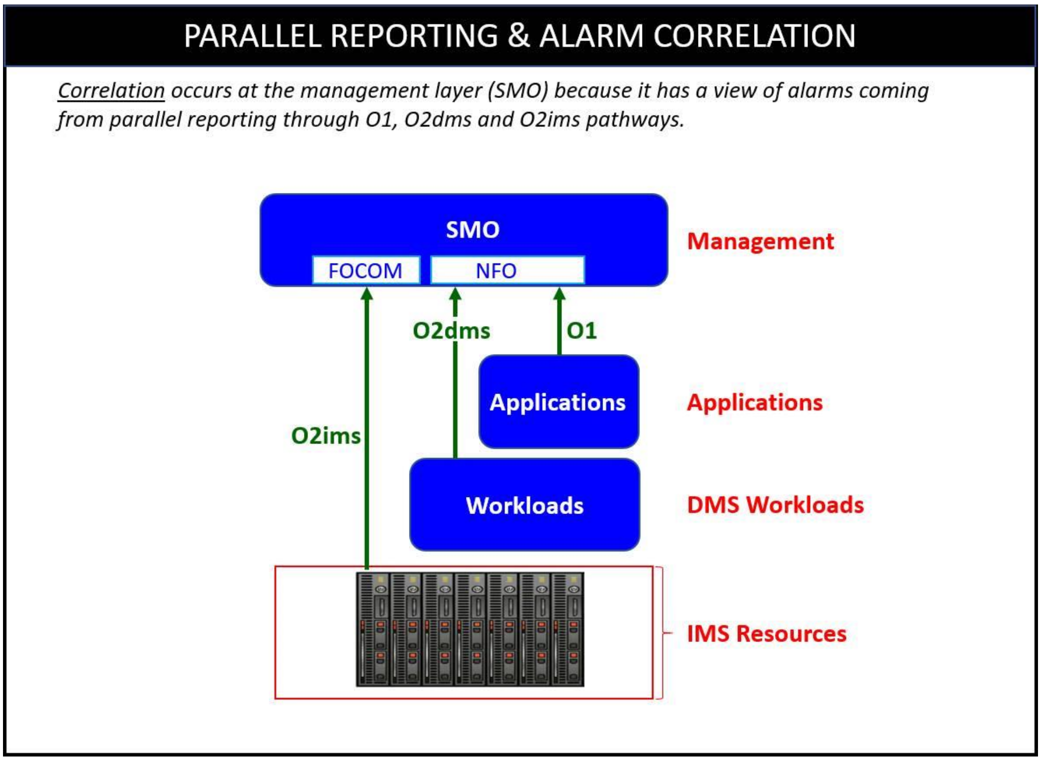

PARALLEL REPORTING & ALARM CORRELATION
Correlation occurs at the management layer (SMO) because it has a view of alarms coming from parallel reporting through O1, O2dms and O2ims pathways.
SMO
FOCOM
NFO
O2dms
O1
O2ims
Applications
Workloads
Management
Applications
DMS Workloads
IMS Resources

Figure 0-1 Parallel Reporting & Alarm Correlation

3.8.13 Escalation Strategy

Fault conditions in an IMS resource may trigger an escalation strategy. An escalation strategy is a mechanism to increase visibility to a condition of duress which is typically a precursor to a more severe situation.

For example, recurring alarms, fault signatures and severity levels may have an escalation strategy to raise higher severity alarms for higher tier recovery actions. As another example, a minor alarm in an O-Cloud resource with a certain occurrence rate within a time window might trigger an escalation strategy to raise a major alarm.

3.8.14 Fault Domains

A fault domain is a "*unit of failure*". A fault domain associates a collection of IMS O-Cloud resources together. We expect a fault domain to be within a Data Center Zone (see the provisioning chapter basic concepts for more detail). Faults could be associated with a fault domain. In the context of IMS fault area, fault domains are an attribute that could be correlative but do not have intrinsic fault operations, they are not a manageable entity.

This would allow operators to design systems that are fault tolerant across fault domains. This might also guide placement strategies, and update cycles to improve reliability and availability.

For example, a network cloud can host some number of servers per rack (e.g., 40 servers) with a number of racks of equipment (e.g., 10 racks) in a single deployment (totaling 400 servers). A fault domain might be defined to be rack. If a fault occurred, this concept would allow an operator trying to isolate the problem down to a rack (the fault domain). It would also allow a user to correlate a problem if multiple racks served by a common infrastructure resource.

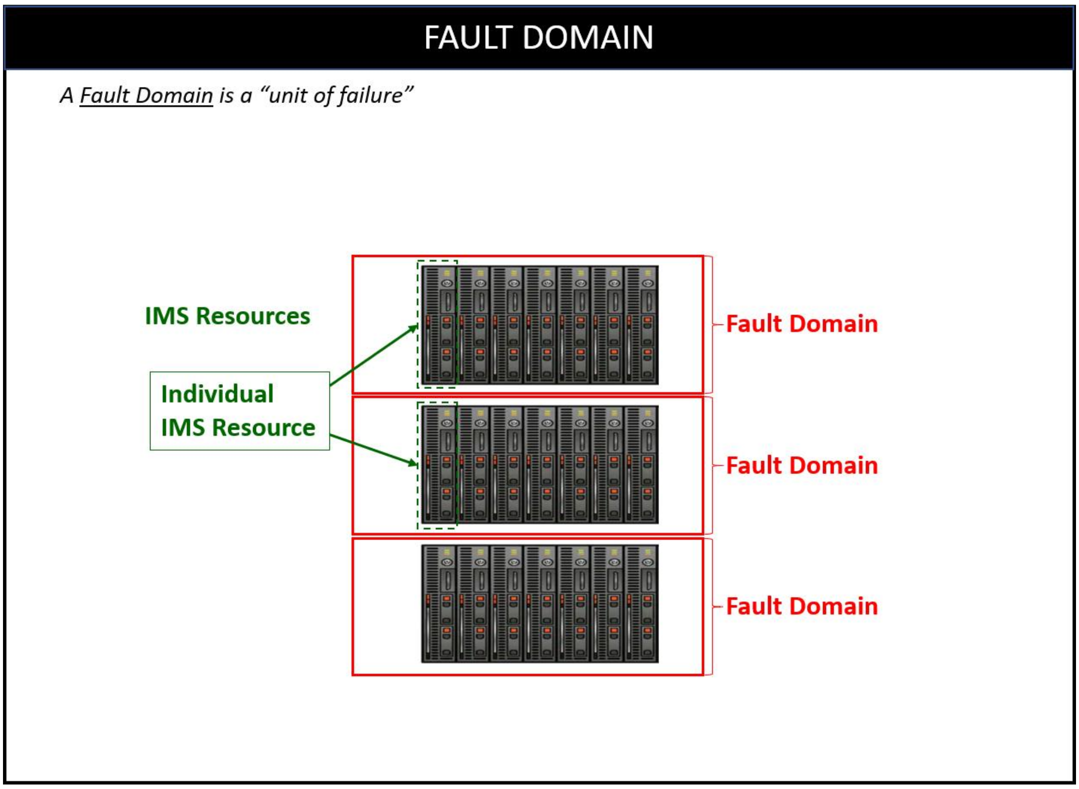


Figure 0-2 Concept of Fault Domains

3.8.15 Fault Tolerance

A fault tolerant system is one that is designed to withstand failures and continue to be able to provide service. Based on observations, some *actions* can be taken to insure reliability. Fault tolerant solutions & strategies will vary between O-Cloud providers and network function providers.

An example of a fault tolerant strategy is the use of geo-redundancy and Data Center Zones. Geographically separating Data Center Zones allows service to continue should a physical natural disaster strike one data center. An *action* might be to migrate traffic from one zone to another.

3.8.16 Metrics

Metrics allow users to define thresholds to monitor and manage IMS O-Cloud resources. From these metrics, a user/system could take action based on threshold crossings and/or chart the metrics over time. Also, to achieve fault tolerance, there could be a mechanism for alerting, monitoring & auditing as a health check.

For example, in the basic fault concept of *Escalation Strategies,* we might need metrics of how **often** a fault occurs within a certain period of time. These *metrics* might be used as a threshold in an escalation strategy for better reliability. If the fault occurred 100 times in an hour action might be taken.

3.8.17 Alarm Suppression

Alarm floods are defined as a multitude of alarm notifications within a short period of time. Alarm floods could be caused by disasters or maintenance work. Thus, it is desirable to be able to suppress alarms in order to avoid alarm floods which distract and overload operators and their monitoring of management systems. The suppression of alarms would cause certain alarms not to be raised when a certain failure(s) happens. The functionality can be managed via the O2ims and/or at the SMO. Furthermore, an operator might be interested in suppressing only certain alarms or types of alarms; hence, an operator could configure the filter for the suppression of alarms via the O2ims.

It is expected that the fault originating from the source of truth, the O-Cloud Resource is still raised to the IMS. It is expected that the O-Cloud Resource faults would still be logged even if they are suppressed. The IMS would also perform the transformation of the fault into an alarm (as normal), and record that an alarm occurred. However, the IMS would suppress the alarm notification over O2ims interface towards the SMO.

There are two motivations for alarm suppression:

1. Reactive suppression – reacting to a situation/disaster which caused/causing an alarm flood.
2. Proactive suppression – a priori, the operator is aware of activity which will cause an alarm flood.

There are two types of suppression:

1. Suppression Type #1 – A mass suppression of alarms, where the suppression filter is requesting for many alarms to be suppressed. This will eventually cause mismatches between IMS and the alarm history at the SMO which would be reconciled later.
2. Suppression Type #2 – The suppression of repeated alarms would suppress an alarm that is being raised many times.

SMO dashboard can filter alarms, which can help a user comprehend and digest a large quantity of alarm information.

- SMO – The SMO Dashboard can choose to filter the display of alarms to make the management of a system more tenable for an operator. Filtering SMO would offer capabilities of filter, the type of alarms that are of interest to an operator.

Alarm suppression occurs in two areas:

1. IMS – The IMS is responsible for transforming faults received from O-Cloud resource into alarms. It would suppress alarms based on its alarm suppression filter.
2. Source of Truth – The O-Cloud resource could suppress faults. This would be implementation specific.

NOTE 1: It is expected that all faults originated from O-Cloud resources would be sent to the IMS.

NOTE 2: It is expected that if the O-Cloud resource suppresses its faults that it would notify the IMS, though this is implementation specific.

Aspects of alarm suppression at the IMS:

1. IMS – Suppresses alarms based on a suppression filter. Suppressed alarms are not raised that would have normally been raised as alarm notifications towards the SMO (client) over the O2ims interface. Consequently, the state of the active alarms in the SMO (client) may be mismatched with the state of alarms in the IMS (O-Cloud) because of alarm suppression.
2. IMS – Respond with its Alarm List information based on the Alarm List query criteria (e.g., a time window which encompasses the suppression period) for purposes of synchronization & reconciliation by a client. The requested

Alarm List can be used by a client to align with its view of alarms after the situation has passed. See the Alarm List Management Use Case for more details.

3. IMS – The IMS would keep an alarm suppression filter(s). The suppression filter is configured by the SMO through the O2ims interface. The IMS may also configure an alarm suppression filter autonomously and inform the SMO through the O2ims interface. The IMS could respond with the current filter that it has, even if it currently has no filter. See the Alarm Suppression Use Case for more details.

4. IMS – IMS would still continue to perform fault-to-alarm transformation and record an alarm occurred irrespective of any alarm suppression filter. This is important that the O-Cloud continue to operate normally so that later when the condition passes the active alarm list can be reconciled by a client.

5. IMS – The cessation of alarm suppression could occur from one of three mechanisms. Alarm suppression can end through specification in the alarm suppression filter request, through a time-out mechanism, and/or by a configurable policy (which would be implementation specific). See the Alarm Suppression Use Case for more details.

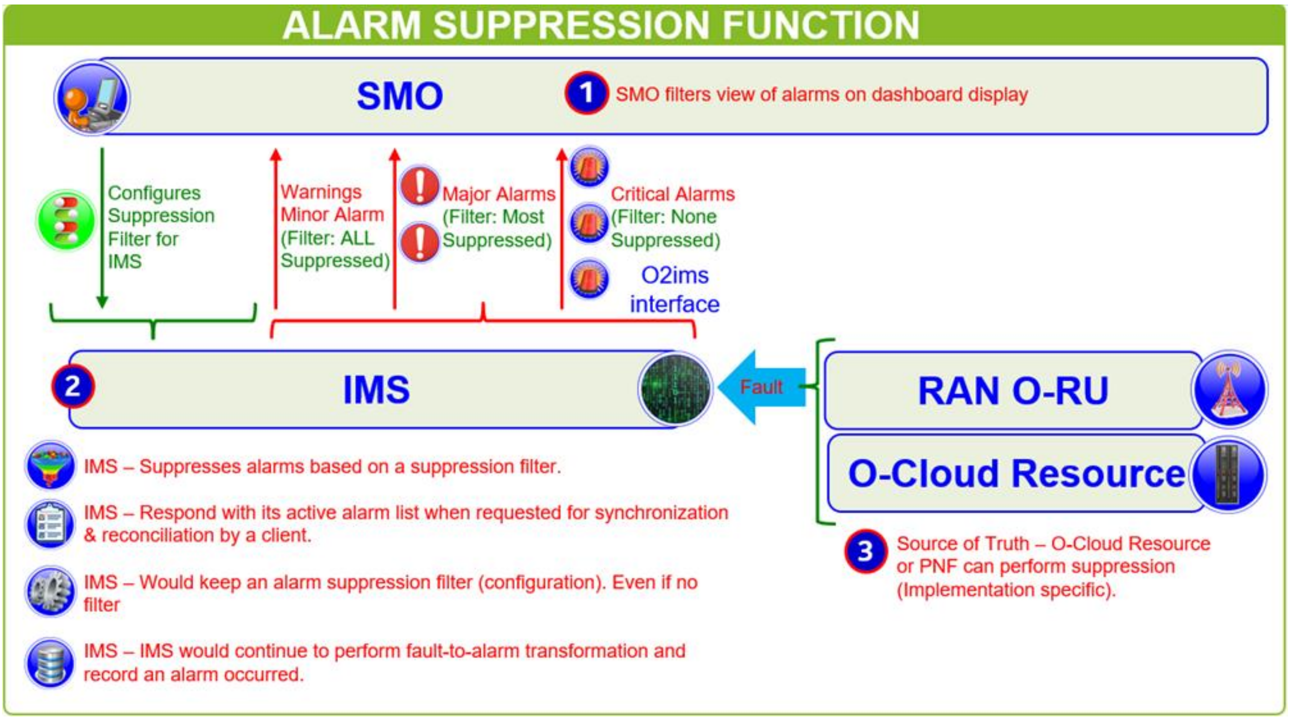

ALARM SUPPRESSION FUNCTION
SMO
1 SMO filters view of alarms on dashboard display
Configures Suppression Filter for IMS
Warnings Minor Alarm (Filter: ALL Suppressed)
Major Alarms (Filter: Most Suppressed)
Critical Alarms (Filter: None Suppressed)
O2ims interface
2
IMS
Fault
RAN O-RU
O-Cloud Resource
IMS – Suppresses alarms based on a suppression filter.
IMS – Respond with its active alarm list when requested for synchronization & reconciliation by a client.
IMS – Would keep an alarm suppression filter (configuration). Even if no filter
IMS – IMS would continue to perform fault-to-alarm transformation and record an alarm occurred.
3 Source of Truth – O-Cloud Resource or PNF can perform suppression (Implementation specific).

Figure 0-3 Alarm Suppression Function

An example of Alarm suppression is as follows. Suppose O-Cloud Resource Compute Node #1 raises a fault #150 towards the IMS. The IMS would change the fault #150 to a corresponding alarm. If a suppression filter were active, causing that alarm to be suppressed, the IMS would not send the alarm notification (see the alarm notification Use Case). Alarm is still in IMS alarm list (not acked). Expected to have a sync, and SMO reconcile alarm.

Some examples of Use Cases for Alarm Suppression are:

- Use Case #1: Something in Maintenance the alarms should be suppressed
- Use Case #2: Disaster turn off all (major, minor, warning) alarms

3.8.18 Fault Supervision

Fault Supervision is a Fault Management (FM) concept regarding the process or operations involved in the fault management of O-Cloud resources.

Examples of Fault Supervision are the operations of Alarm Acknowledge and Alarm Clear. See the corresponding use cases in the Orchestration and Cloudification Use Cases and Requirements for O-RAN Virtualized RAN [13].

3.8.19 Logs for Fault Management

There are three basic kinds of logs that might be kept within the Cloud (both exposed O-Cloud and Cloud infrastructure):

- ALARM LOGS – Alarm logs are a record of the alarms that have been raised by the IMS. The IMS keeps and produces an Alarm Log. It is expected that the IMS logs every alarm (including suppressed alarms) that it sends to the FOCOM (SMO). Alarm Logs are exposed in an O-Cloud to northbound entities.

NOTE 1: It is also possible to log additional events or actions against alarms, such as clearing and acknowledgment of alarms. See the Log Query and Logging Management Use Cases for further details [13].

- FAULT LOGS – All detected Faults are raised by an O-Cloud Resource. Faults information is stored by the O-Cloud Resource, either remotely or locally, in a Fault Log. Not every fault necessarily raises an alarm, hence the necessity to log events for both fault information and alarms. The present document does not specify or describe the creation and operation of fault logs. Disclaimer: it is possible to combine Fault and Debug logs together. Fault logs, once activated, would be exposed to northbound entities. Fault Logging may be part of the general event logging and is a concept where fault information is stored by the O-Cloud Resource (the source of truth), either remotely or locally, for a configured amount of time. The concept of Fault Logging and storing fault information allows a client, algorithm, or user to query the fault log and analyze those faults. Faults that are intermittent, repetitive, or transient can be identified this way. For example, Debug traces used by some O-Cloud providers might be a way to facilitate fault logging meaning that certain condition can trigger writing to the fault log.

- DEBUG LOGS – The present document does not specify or describe the creation and operation of Debug Logs. They might be able to be turned off. The log level would typically be configurable (at the source, middle, and end user level). For example, a typical implementation may use syslog (defined in RFC 5424, see reference [17]) which defines a “log level” to select informational (most detail), trace, warnings, errors, to critical (least detail) messages. Disclaimer: It might be possible that Fault logs and Debug logs might be combined. Debug logs, once activated, would be exposed to northbound entities.

A Timestamp is a particular point in time when an entry in a log is created, and timestamps are typically used for correlated analysis across logs and in maintenance activities. Therefore, all log files typically follow a timestamp format which would enable for correlation across log files.

Logging Management uses mechanism(s) such as Retention Period to manage the length of time that entries in a log are maintained. For example, logs for an O-Cloud Resource could be stored for 1 week.

There may be other logs besides Alarm, Fault and Debug logs that might be kept in the O-Cloud. The present document does not specify or describe what other types of logs beyond Alarm, Fault and Debug logs are to be handled by the O-Cloud.

The following diagram illustrates the basic concept for Fault and Alarm Logging:

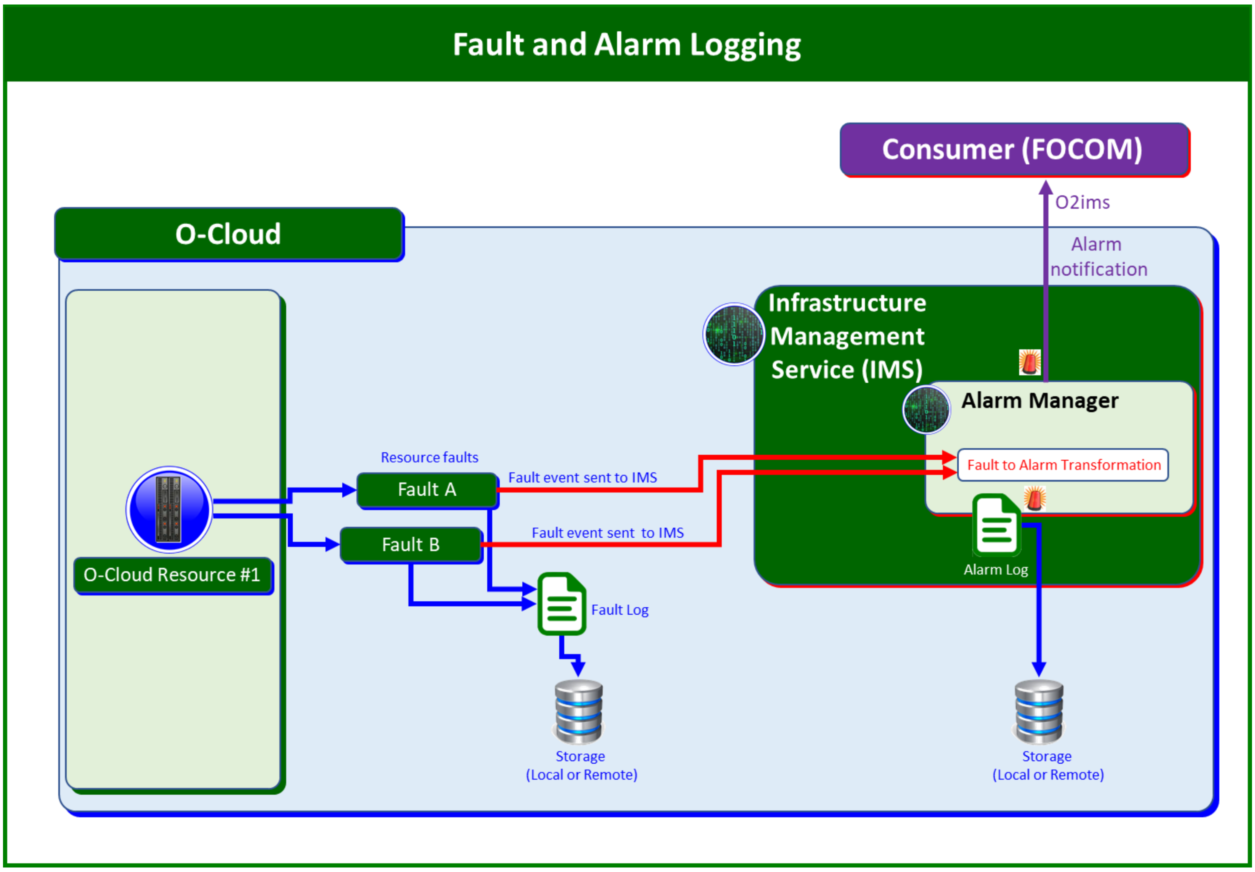


Figure 0-4 Fault and Alarm Logging

In the above diagram, O-Cloud Resource #1 encounters two faults, Fault A and Fault B. It sends those as Fault events to the IMS when they occur. O-Cloud Resource #1 might also store the occurrence of Fault A and Fault B into a Fault Log in local or remote storage. The IMS would assess and transform Fault A and Fault B into the appropriate Alarm(s). Suppose in this case, the IMS determines that Fault A would not need to be raised as an Alarm; however, the IMS determines that Fault B needs to be transformed into Alarm#10. The IMS would store the Alarm#10 into its alarm log to be stored either locally or remotely. The IMS would raise the Alarm#10 as an alarm notification over the O2ims interface.

NOTE 2: The Alarm Manager shown in the diagram above is software which transforms Faults to Alarms that need to be exposed. The alarm manager depicted in figure 1 is only for illustrative purposes to show the process of logging alarm done by the IMS.

3.8.20 Logging Level

The Logging Levels adjust the amount of detail recorded in Fault and Debug Logs captured from O-Cloud resources. The Logging Level is associated with the severity levels and thus what is captured in logs.

The motivation for the Logging Level concept is to change how much information should be captured and sent from O-Cloud resources. For example, while debugging, a developer might raise the Logging Level to see more faults.

NOTE: Logging Levels may apply differently to Fault Logs vs Debug Logs.

The notion of Logging Level is applicable to the source generating logs and from the perspective of the viewer. At the generation side, the Logging Level is adjusted to change the amount of data captured and sent; at the viewing side, the Logging Level indicates the level of detail recorded by logs. The higher the Logging Level, the more data that is captured and potentially sent on the generation side.

Some relevant specifications include IETF RFC 5424 [17], The Syslog Protocol. ETSI GS NFV-SOL 009 v4.3.1 clause 8 which is the API on Log Management [18]. These describe how a consumer can set up a Log job. Log jobs collect logging information based on criteria with the possibility to indicate Logging Level. The SOL 009 specification: (1) describes the management of NFV-MANO, (2) It has a Log management API, and (3) the Log management API refers to IETF RFC 5424.

3.8.21 Alarm List

The Alarm List in the O-Cloud contains the history of Alarm Events that have been detected by the IMS. Each IMS manages one and only one Alarm List. The entries in the Alarm List are Alarm Event Records against resources and their respective resource type.

The Alarm List contains both the active and inactive (a.k.a. cleared) alarms.

The Alarm List concept is used in the Alarm List Management and Alarm Purge Use Case in the O-RAN.WG6.Orch-Use-Cases [13], Clauses 3.7.11 and 3.7.13 respectively.

Alarm Life Cycle Management is described in IETF RFC 8632 [20]. 3GPP TS32.111 [21] introduces concepts related to Alarm List.

NOTE: The present document considers that both acknowledged and cleared alarms are in the Alarm List. This differs from 3GPP TS32.111 [20] where clearing and acknowledging alarms causes them to be removed from the Alarm List.

3.8.22 Retention Period

Alarm Event records in the Alarm List conform to an Alarm Management Policy which specifies how long Alarms should be kept in the Alarm List.

When Alarm Event records in the Alarm List exceed the time period for retention, those records are purged or archived based on Alarm Management Policy.

The Retention Period is not set per resource type. Instead, the Retention Period is applicable to all entries in the Alarm List. Therefore, an Alarm List has one and only one Retention Period attribute.

The Retention Period applies to alarms that are both inactive and acknowledged.

3.8.23 Alarm Purge

The concept of Alarm Purge is defined by the operation of removing Alarm Event Records from the Alarm List permanently. The Alarm Event records to be purged are determined by the Retention Period. See the concept from Clause 3.8.22. Alarms in the Alarm List that reach the age of the Retention Period will be purged. When an alarm in the Alarm List is older than the Retention Period, the IMS automatically purges those old Alarm Event records if they are also inactive and acknowledged alarms. Alarms can also be manually purged through an Alarm Purge request from an O-Cloud operator. In this case, alarms to be purged are removed from the Alarm List irrespective of the Retention Period.

For further information, see IETF RFC 8632 [20].

NOTE 1: Only inactive alarms and acknowledged alarms can be purged.

NOTE 2: A forced Purge of alarms might be used if the storage capacity for the Alarm List approaches or is at capacity; and an operator wants to remove (e.g., warning, and minor) alarms from the Alarm List to free up space.

NOTE 3: If an alarm condition persists after the original alarm is purged, it will reappear as a new alarm (spurred from that same condition) back in the Alarm List.

3.8.24 Archiving Alarms

The concept of Archiving Alarms is defined by the operation of moving Alarm Event Records to a long-term storage or collection area.

NOTE: The Archiving of Alarms may mean moving them off the O-Cloud to a separate off-site storage location. Archiving alarms is likely to be handled by tiered storage paradigms which will handle the archival of alarms. This would be implementation specific.

3.9 O-Cloud Performance Basic Concepts

3.9.0 Overview

The following section describes basic concepts for O2 Performance Concepts.

In general, the purpose of performance is to report operational information related to O-cloud resources. Typically, performance information allows an operator or administrator of the O-cloud a sense of how well the system is operating. It is distinct from faults in that it is not about failures in the system but about how well the overall system is performing. Though, faults or alarms may negatively impact performance of the O-Cloud which might be observable in performance measurements.

Performance measurements are typically captured periodically through time. They are collected and stored at regular intervals by the system in order to gauge the performance of a system over a period of time. This allows for analytical operations to be performed on the collected data and statistics to be built over time. This can tell an operator or system analyst whether they have sufficient capacity in a network based on the demands of the network. This can be vital for making business operational decisions such as scaling a network.

The following are basic performance related concepts that will be used in the O2 Cloud performance use cases and modeling.

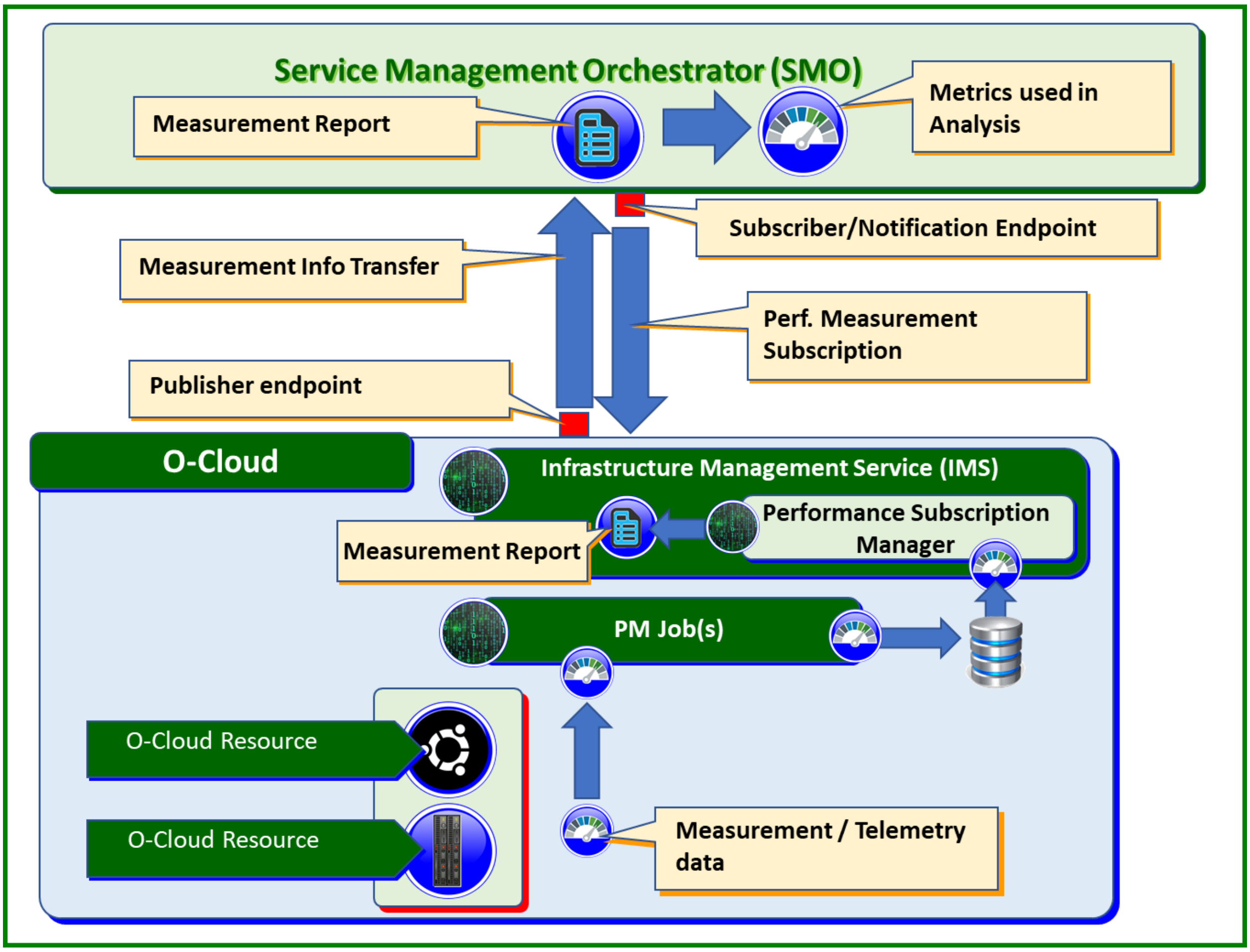

Service Management Orchestrator (SMO)
Measurement Report
Metrics used in Analysis
Subscriber/Notification Endpoint
Measurement Info Transfer
Perf. Measurement Subscription
Publisher endpoint
O-Cloud
Infrastructure Management Service (IMS)
Performance Subscription Manager
Measurement Report
PM Job(s)
O-Cloud Resource
O-Cloud Resource
Measurement / Telemetry data

Figure 0-1 Performance Measurement Reporting

Figure 3.9-1 shows the basic flow of Performance Data framework. Measurement and Telemetry data is collected from O-Cloud resources by PM Job(s) and stored locally. The Performance Subscription Manager (PSM) within the IMS can retrieve the data to be sent to the SMO. There is a consumer (subscriber) notification end point that is used to send the performance data to. The notification end point is known to the PSM from the subscription. Measurement reports are composed by the PSM and sent to the SMO. The SMO can then be used to perform analysis on the Measurement Report.

Performance measurements in an O-Cloud might streamed (pushed) or it might be pulled (measurement reporting files) by a user or a PM job in the SMO (consumer). For performance measurements to be streamed, a subscription needs to be created.

The subscription and criteria indicate an end point for where the measurements are sent to. The notification end point can be updated by performance subscription update.

The subscription will indicate the interval. A capabilities exchange could occur between the SMO and the O-Cloud indicating the mechanism & formats supported by the O-Cloud. The SMO can then subscribe to the mechanisms & formats within the subscription filter. In the capabilities exchange, the IMS would report the jobs and their collection intervals. A performance subscription can be created, queried, updated, and deleted over O2ims.

3.9.1 O-Cloud Performance Measurements

Measurements are defined as the act of quantitatively checking the O-Cloud resources for defined gauges of operation. A measurement is an individual quantitative gauge of a quality.

For example, the O-Cloud might measure the computational performance of an O-Cloud compute resource in terms of CPU utilization.

3.9.2 Performance Indicators

Performance indicators take execution measurements and use them in a way that can give a meaningful notion of how a system is utilizing its resources and/or how well a system is operating. Indicators might be local in the sense of a function or more broader in scope for a group of functions. Performance indicators are typically a formula-driven application usage of performance measurements. Thus, the indicators are amenable to statistical analysis. The indicator formula needs to be exposed so that users can understand the meaning of the indicator. Indicators use a measurement, or gauge, to show an aspect of a thing. Key Performance Indicators (KPI) are performance indicators that are typically used to detect and report overall network health.

For example, the CPU utilization might be a performance measurement. A performance indicator might be the fraction of time that a system is over a threshold, such as 80% CPU utilization. This would be a simple ratio calculation that could give a sense of how close to being overloaded a system is.

3.9.3 Performance Metric

A performance metric represents standards of measurements. Metrics might be pre-defined, or they might define performance threshold targets to try to achieve. Metrics may utilize measurements and indicators. Performance metrics could be used by O-Cloud operators to achieve OPEX business objectives.

For example, a performance metric could be a target that an operator would like to have a system operate at. For example, a metric to operate at five-9's reliability, or CPU utilization that is under a threshold.

3.9.4 Performance Reports

Performance reports are a collection of measurements. Usually, they are gathered in a file; however, they may also be sent as a data stream. Typically, reports are created at regular (configurable) intervals or triggered by thresholds. The creation of performance reports may also be manually invoked.

For example, The O-Cloud might compose a performance report which is comprised of several performance measurements recorded and stored at regular intervals.

3.9.5 Performance Management (PM) Job

3.9.5.1 Basic PM Job Concepts

A performance management job is a task whose purpose is to collect measurements, creating a measurement file, and get a performance report according to criteria set by the user. The criteria could be used to specify measurements to collect.

NOTE: 3GPP standards also discuss measurement jobs see references, 3GPP TS28.550 [14], 3GPP TS28.531 [15], ETSI SOL003 [16].

For example, a PM job might be scheduled to run at 11 PM with the criteria to retrieve CPU measurements for O-Cloud resources and compose them into a measurement report.

PM job can be created, queried, updated, suspended, resumed, and deleted over O2ims.

3.9.5.2 PM Job States and Statuses

A PM Job can have one of the following states. Each State is mutually exclusive from the others.

- ACTIVE – An Active PM Job is one that is operational and currently collecting one or more measurements.
- SUSPENDED – A Suspended PM Job is one that a consumer has issued an operation to stop the collection of all measurements. A Suspended PM Job when resumed goes back to the Active state. A delete operation can only be processed for a Suspended PM Job.
- DEPRECATED – A Deprecated PM Job is one that has had a delete operation issued for it; and is now pending delete. It is waiting on a pending purge of all the measurements collected by the PM Job from the PM Store whereupon the PM Job will be deleted. The system purges measurement(s) that are older than a configurable retention period. Measurements previously collected by the deleted PM Job could still be requested which would cause referential integrity issues if that PM Job were to be deleted which is why the *Deprecated* state and *Pending Delete* status exist.

A PM Job can have one of the following statuses. Each Status is mutually exclusive from the others.

- RUNNING – A Running status applies to an Active PM Job that is currently performing metric collection without issue.
- FAILED – A Failed status applies to an Active PM Job that is failed and is no longer collecting any measurements.
- DEGRADED – A Degraded status applies to an Active PM Job that is experiencing one or more issues but has not fully failed. A Degraded PM Job could still be collecting all or some subset of its measurements.
- IDLE – An Idle status applies to a Suspended PM Job; and is therefore, no longer collecting any measurements.
- PENDING DELETE – A Pending Delete status applies to a Deprecated PM Job which remains in the Deprecated state until all measurements associated with it have been purged. Once all its measurements have been purged, the PM Job is removed from the system.

3.9.5.3 Preinstalled PM Job

A Preinstalled PM Job is a type of PM Job that is created automatically by the O-Cloud. Preinstalled PM Jobs can be created during O-Cloud genesis (O-Cloud start up). Preinstalled PM Jobs behave like other PM Jobs. One of the purposes of the measurements collected by Preinstalled PM Jobs is for internal O-Cloud PM operation and management.

An IMS consumer cannot change or alter a Preinstalled PM Job. Preinstalled PM Jobs may be deleted by an IMS software update. The SMO or other entity can subscribe to Preinstalled PM Jobs and their attendant measurements.

3.9.6 Source of Truth

The source of truth is the arbiter of discrepancies in metrics. In an O-Cloud IMS Performance context, it is expected that the "*source of truth*" will be the O-Cloud (IMS + O-Cloud infrastructure resources). The PM job manages the transformation of data into a performance report. The PM job requests data from the O-Cloud (the Source of Truth) to build the performance report.

For example, suppose there is a metric from the source of truth with a reported value of 50, and the performance report has a value of 20. That value of 20 might have been a result of corruption in sending, composition, or a software bug. In order to obtain the proper value, the source of truth can be consulted.

3.9.7 Performance Threshold

A performance threshold is a level or value which is used to check if a measurement exceeds or stays within. Typically, thresholds are used to trigger behavior or policies in a system.

For example, a user might set a CPU utilization threshold of 80%.

3.9.8 Performance Dictionary

A Performance Dictionary is a reference which details information on performance metrics or KPIs that are relevant to one or more O-Cloud Resource(s) and that would be reported by the IMS in the O-Cloud so that the SMO could act on those measurements. The Performance Dictionary would be onboarded when a new O-Cloud Resource Type is added, for example during the process of 1) O-Cloud registration, 2) IMS Software update 3) O-Cloud Resource addition and 4) O-Cloud Resource software update; and that O-Cloud Resource Type would come with an associated a Performance Dictionary. The Performance Dictionary for an O-Cloud Resource includes all measurements supported by the O-Cloud Resource Type including ones defined by standards and the reference to the standard.

For example, there will likely be some standardized measurements. For a CPU, that might be CPU utilization measured in percentage. However, there could also be a vendor's CPU that might have additional features such as pinning or huge page functionality.

3.9.9 Performance Measurement Operation Example

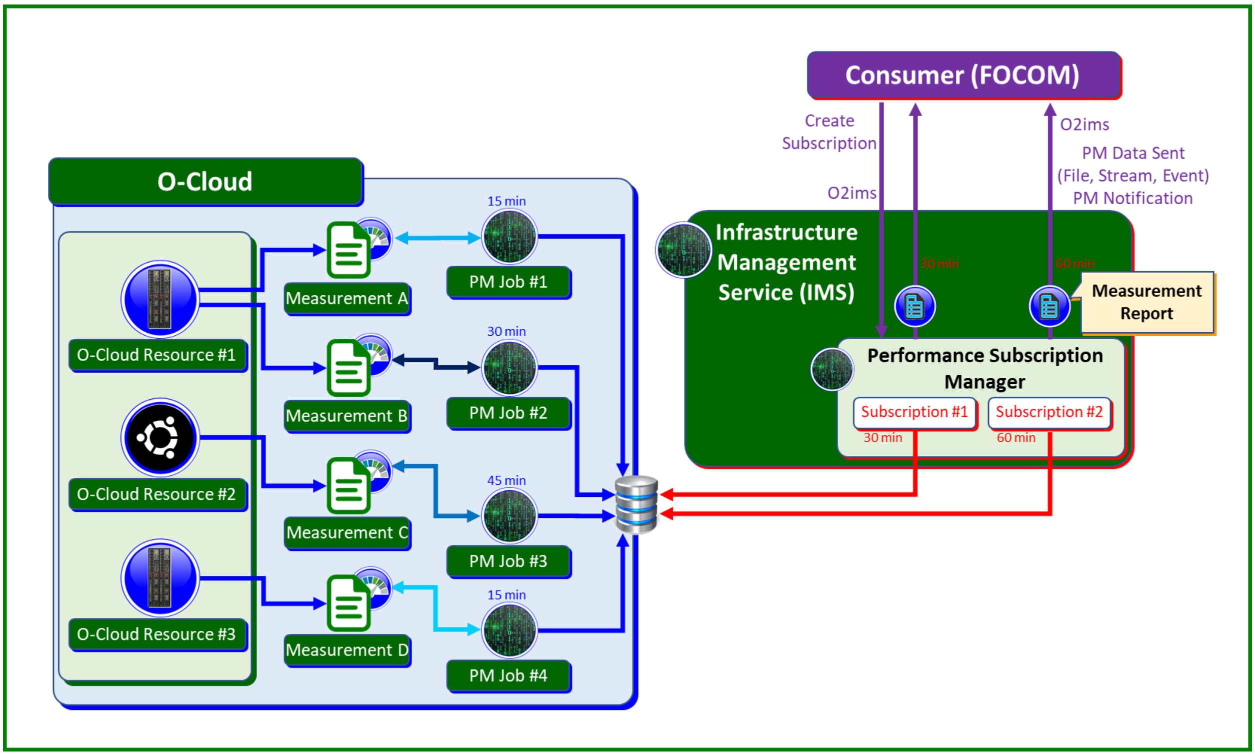


Figure 0-2 Performance Measurement Operation Example

PM Job is a software task that collects measurements or telemetry data from O-Cloud resource(s). The PM Job primary function is to collect measurements or telemetry data based on a schedule; and it writes its measurements to a local repository. The PSM may take the raw data from local storage of data collected from PM Jobs and transforms it from the format in the local store to a report format. For example, the PM Job #1 collects as Measurement A values 1, 3, 8. That is stored locally. The PSM takes those values to compose a performance report shown in figure 3.9-2. The blue shaded arrows are related to the PM job data collection. The red colored arrows are related to the PSM. The purple arrows are activities that happen over the O2ims interface.

Figure 3.9-2 illustrates an example of the Performance Management framework. In this diagram, there are three O-Cloud Resources identified as Resource #1, Resource #2, Resource #3. There are four measurements identified as Measurement-A, Measurement-B, Measurement-C, Measurement-D that originate from these three different O-Cloud Resources. In this example, Measurement-A and Measurement-B originate from Resource #1. Measurement-C originates from Resource #2 and Measurement-D comes from Resource #3. There are four PM Jobs shown that collect Measurements specified when they are created. PM Job #1 collects Measurement-A with a collection interval of every 15 minutes. PM Job #2 collects Measurement-B with a collection interval of every 30 minutes. PM Job #3 collects Measurement-C with a collection interval of every 45 minutes. PM Job #4 collects Measurement-D with a collection interval of every 15 minutes.

NOTE: If an operator would like to adjust the collection interval of a PM Job or adjust what measurements that PM Job is collecting it would require a PM Job update (See the PM Job Update Use Case [13].

The PSM within the IMS retrieves the data and composes a Measurement Report to be sent to the SMO. It does so at the specified reporting interval as defined by the subscription. For example, figure 3.9-2 shows two subscriptions. Subscription #1 has a reporting interval of 30 minutes with some combinations of measures that it reports on. Subscription #2 has a reporting interval of 60 minutes.

The data store, shown as (🛢) in figure 3.9-2 is an internal store in the O-Cloud that is only accessible by the O-Cloud (in this case the PM Jobs and the PSM). The PM Job(s) may also direct O-Cloud Resource(s) to directly write primitive measure(s) to a data store. The realization of the PM Jobs is left up to implementation.

3.9.10 Performance Measurement Transformation

Performance Measurement Transformation is defined as taking primitive performance measurements and consequently adjusting those collected primitive measurements either syntactically or semantically. There are two kinds of performance measurement transformations:

- Syntactic performance measurement transformations that change the format of the primitive measurement for storage and later retrieval. An example of syntactic performance measurement transformation might take raw data from a O-Cloud resource and store it into a JSON or Yang format as necessary.
- Semantic performance measurement transformations occur through meaningful calculations of primitive measures. An example of semantic performance measurement transformation are averages, means, and peaks that are derived from primitive measures.

Transforming data is used during Performance Measurement reporting process. The PM Job creation would specify the measures and transformations their inputs, and outputs.

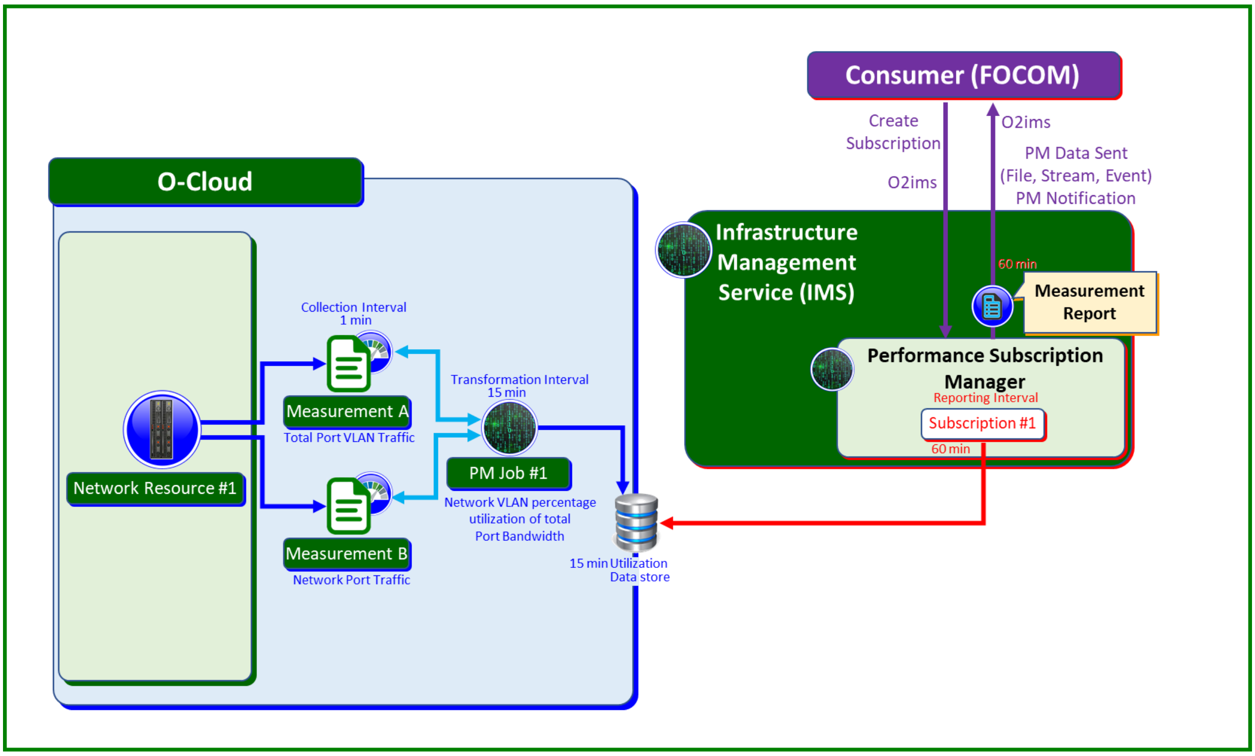

Consumer (FOCOM)
Create Subscription
O2ims
O2ims
PM Data Sent (File, Stream, Event) PM Notification
O-Cloud
Infrastructure Management Service (IMS)
60 min
Measurement Report
Performance Subscription Manager
Reporting Interval
Subscription #1
60 min
Collection Interval 1 min
Measurement A
Total Port VLAN Traffic
Transformation Interval 15 min
PM Job #1
Network VLAN percentage utilization of total Port Bandwidth
Network Resource #1
Measurement B
Network Port Traffic
15 min Utilization Data store

Figure 0-3 Semantic Performance Measurement Transformation Example

An example of PM data collection using transformation is shown in figure 3.9-3, O-Cloud Network Resource #1 produces two primitive measures every minute. Measurement-A is the Network Port Traffic and Measurement-B is the Network Port VLAN Traffic. PM Job #1 collects these two primitive measures and performs semantical performance measurement transformation to produce Network VLAN percentage utilization of total port bandwidth which is Measurement A divided by Network Port Total Bandwidth. Another transformation would be Measurement B divided by Network Port Total Bandwidth because the utilization is defined as demand divided by capacity. PM Job #1 does this computation at a transformation interval of 15 minutes and writes the utilization to a local data store. Subscription #1 was created to report on the utilization. Its reporting interval is 60 minutes. Thus, it would have four data points for utilization every hour because PM Job #1 would store four data points within an hour. The Subscription #1 does not perform any further calculations.

Note that PM Job #1 could also collect other primitive measures as well, and directly write that into the data store. That is, PM Job #1 is not always doing a semantical calculation or performance measurement transformation on data.

3.9.11 Retention Period

Performance Measurement records in the Performance Measurement Store conform to a Performance Management Policy which specifies how long Measurements should be retained in the Performance Measurement Store.

Performance Measurement records are retained according to the retention policy of the O-Cloud based on the configured retention period.

The Performance Measurement Store has one and only one Retention Period attribute, and it applies to all entries in the Performance Measurement Store. When Performance Measurement records in the Performance Measurement Store exceed the time for retention, those records are deleted or archived.

There is a relationship between a deprecated Performance Measurement Job and the retention period. A deprecated Performance Measurement Job cannot be purged until all its associated Performance Measurement records in the Performance Measurement Store are deleted due to being older than the retention period.

3.10 O-Cloud Operational Modes

3.10.1 Overview

Section 3.10 describes basic concepts related to the operational modes of O-Cloud Nodes.

The state and status of the O-Cloud Nodes can be managed by SMO via O2 interfaces. In the following, the combination of the state and status of O-Cloud Nodes managed by SMO and the required operations specific to each mode are referred generically as operational modes. The following operational modes are further described: Maintenance mode (see section 3.10.2) and Test mode (see section 3.10.3).

The O-Cloud operator can switch the operational mode of O-Cloud Nodes to handle specific operational requirements and events on the O-Cloud Nodes, for instance, a planned maintenance and/or functional test after fault recovery. In these cases, the behaviour and usage of O-Cloud Nodes changes according to the purposes of the operational mode.

ITU-T X.731 defines administrative state and availability/control status which serves as the foundation of the maintenance mode and test mode concepts discussed here.

The state transitions, what are legal states, and moving from one state to another will be described in future versions of the present document.

3.10.2 Maintenance Mode

Maintenance Mode is one of the operational modes of O-Cloud Nodes. The purpose of the Maintenance Mode is to indicate that it is under maintenance by changing the state of the managed entity. In Maintenance mode the resource is intended to be isolated and not used. This mode can be used in various use cases for fault recovery, O-Cloud Resources upgrade, and planned maintenance.

The managed objects considered in this mode are the O-Cloud Nodes.

NOTE 1: There could be other O-Cloud managed objects, to be described in future versions of the present document.

O-Cloud operator switches the operational mode of O-Cloud Nodes into the Maintenance Mode using the O2ims and/or O2dms. While the O-Cloud Nodes are in Maintenance Mode, NFs/NF Deployments cannot utilize the O-Cloud Resources.

NOTE 2: Implications of Maintenance Mode to different types of O-Cloud managed objects, the detailed meaning of Maintenance Mode for different forms of deployments (e.g., VM, container) may be described in future versions of the present document.

3.10.3 Test Mode

Test Mode is one of the operational modes of O-Cloud Nodes. The purpose of the Test Mode is to enable the means for the O-Cloud operator to test the O-Cloud Nodes using testing procedures. In Test Mode the resource can still be used but in a controlled manner typically for testing purposes. This Mode can be used after performing a maintenance of the O-Cloud Nodes, or after the O-Cloud initialization, or after installing new individual O-Cloud Nodes in the O-Cloud.

The managed objects considered in this mode are the O-Cloud Nodes.

NOTE 1: There could be other managed objects, but those may be described in future versions of the present document.

O-Cloud operator switches the operational mode of O-Cloud Nodes into the Test Mode using the O2ims. O-Cloud Operator can check that the O-Cloud Node is able to operate correctly from a functional (e.g., behaviour, capabilities) and non-functional (e.g., performance) views. While the O-Cloud Nodes are in Test Mode, only test mode NF Deployments can utilize the O-Cloud Resources.

NOTE 2: Implications of Test Mode to different types of O-Cloud managed objects such as a Test Mode NF Deployment, the detailed meaning of Test Mode for different forms of deployments (e.g., VM, container) may be described in future versions of the present document.

NOTE 3: Alarm handling in Test Mode will be clarified in Alarm Suppression use case.

Annex A (informative): O-Cloud and IMS Initialization

A.1 Overview

The present document does not dictate how O-Cloud is initialized and how initial endpoint for SMO to connect to is done. But here are some logical steps of the process to make it happen. This is included for informational purposes and not part of formal specification, please see Orchestration Use Cases [13].

1. SMO manually populated with a server that will be part of HW that is in O-Cloud. That server has connectivity to all other HW in the O-Cloud currently (at the start). The choice of that initial server is not important and is used only for bootstrapping (Genesis).
2. SMO deploys a minimalistic IMS functionality to initial server so it can do a subset of IMS functionality and bring O-Cloud up. At a minimum that IMS services should include infrastructure discovery, inventory, and administration within the O-Cloud; and deployment of O-Cloud with full IMS and DMS functionality. The configuration of the network might be sent to the server as part of its startup. In which case it might not have connectivity to all other HW but will have the ability to establish connectivity by configuring a transport path to servers in a non-collocated resource pool.
3. After O-Cloud deployment the initial server is no longer needed for minimalistic IMS services and can be returned into a pool of resources under O-Cloud control. The registration of the complete IMS services endpoint in the SMO, via a callback provided to the initial server, exposes the internal cloud services to the SMO via the IMS SBI.

The present document does not define how O-Cloud with IMS services come into being. So, the above steps are provided for informational purposes only.

A.2 UML sequence diagram for O-Cloud and IMS initialization

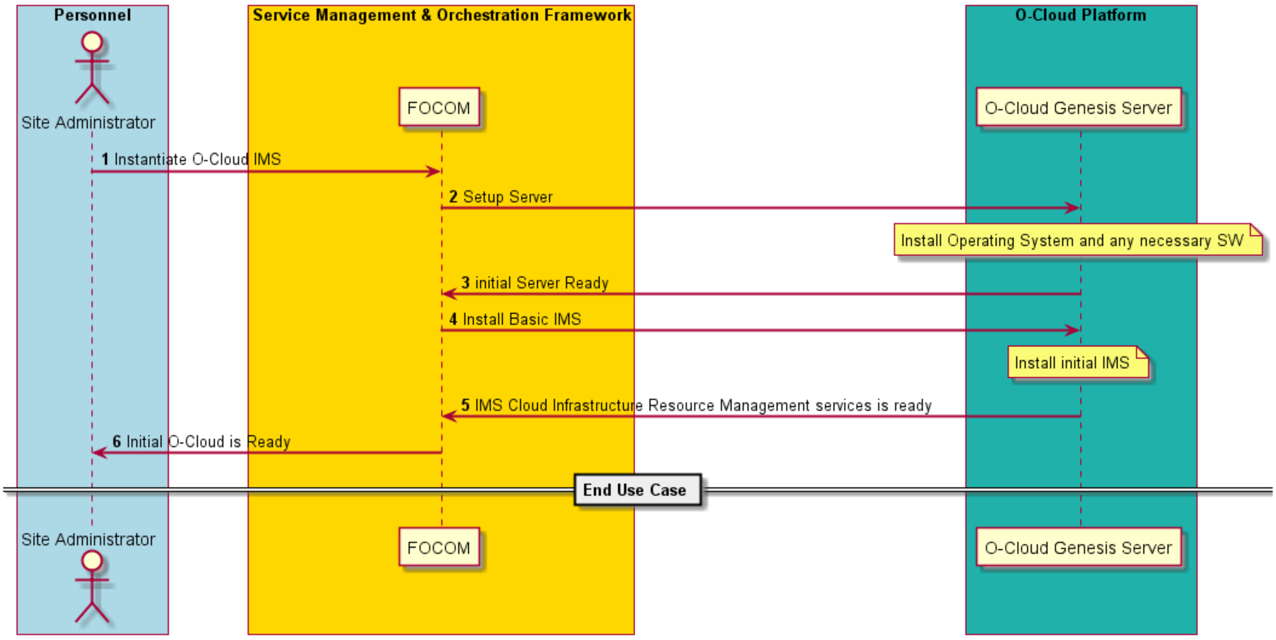

Personnel
Service Management & Orchestration Framework
O-Cloud Platform
Site Administrator
FOCOM
O-Cloud Genesis Server
1 Instantiate O-Cloud IMS
2 Setup Server
Install Operating System and any necessary SW
3 initial Server Ready
4 Install Basic IMS
Install initial IMS
5 IMS Cloud Infrastructure Resource Management services is ready
6 Initial O-Cloud is Ready
End Use Case
Site Administrator
FOCOM
O-Cloud Genesis Server

Figure A-1 O-Cloud and IMS Initialization

Annex B (informative): O-Cloud Resource View examples and operations that use these Views

Below follows a set of realistic examples of the four Resource views as specified in section 3.4.3.3 above.

- Cloud Infrastructure Resources would typically include physical and logical Servers, Switches, Routers and Site Network Fabrics (as seen from the Green Robot head view)
- Abstracted Resources would typically offer physical O-Cloud Computer Systems, logical VMs and O-Cloud Site Networks (as seen from the Yellow Robot head view)
- Assigned Resources would typically include Nodes, Node Clusters and Node Cluster Site Networks (as seen from the Blue Robot head view)
- Consumed Resources would typically be CPU consumption of a Nodes, Packets transported on a NIC and Switch Port, Bandwidth used on a Site Network Fabric and Gateway (as seen from the Red Robot head view)

Below follows a few natural flow examples of Resource handling that includes the four O-Cloud Resource views,

1. A Cloud Site installation staff introduce a new physical Cloud Infrastructure Resource that is started up and found to be healthy and made available for the O-Cloud provisioning system as an O-Cloud ResourceType (as exemplified by the steps below)
   a. IMS adds the newly available Abstracted Resource to an O-Cloud Resource Pool, adds the O-Cloud Resource to the O-Cloud inventory and notify the subscribers of the inventory data of the relevant new/updated content
   b. SMO or IMS (if there is a pending demand for such a new Abstracted Resource) book the Abstracted Resource as a new Assigned Node for a Node Cluster
   c. DMS now have more capacity and possibly also new capabilities through the new Assigned Node and places a NF Deployment on the Node Cluster
   d. The NF Deployment (workload) Consumes the new Resource and gets the relevant metrics through the running Operating System drivers
2. FOCOM requests a new Node Cluster to be fulfilled by IMS without having any pre-assigned O-Cloud Resources available (as exemplified by the steps below)
   a. IMS searches for available and suitable Cloud Infrastructure Resources for the Control Plane Nodes and the Worker Nodes that can fulfil the requested capacity and characteristics of the Node Cluster
   b. IMS books the found and required Cloud Infrastructure Resources and creates Abstracted Resources for the Worker Nodes and the Node Cluster Site Networks that are then assigned to suitable Resource Pools
   c. The Abstracted Resources are Assigned to Nodes for the Worker Nodes and Node Cluster Site Networks for the Node Cluster
   d. IMS now starts up the Node Cluster using the Cluster Infrastructure Resources for the Control Plane Nodes, the Assigned Nodes for the Worker Nodes and the Assigned Node Cluster Site Networks for the Node Cluster communication
   e. IMS sets up the appropriate DMS user credentials and notifies FOCOM about the new inventory content as well as the DMS end-point that has been created
   f. SMO makes the new DMS available to the NFO
   g. NFO starts using the new DMS and request NF Deployments on it
   h. The NF Deployments (workloads) consume the Assigned Node Cluster and Node Resource and gets the relevant metrics through its running Operating System drivers

Annex C (informative): IMS example usage of ICS applied to ResourceTypes

Below follows some examples of the Initiation Configuration Set as specified in section 3.4.3.7 above. These are only exemplifications to better understand the concept. The examples might include some non-aligned terms, and they should not be taken as new concepts or terms.

ResourceTypes that are well suited to use ICS are Compute, HW-Accelerator, Site Network Fabric, storageunits and portgroups on any device e.g., Site Network Fabric and Compute. As an example of this flexibility for the deployed ResourceTypes, it is completely up to a specific Compute ResourceType specification to have its port configurations in the Compute ICS, in a separated portgroup ICS or have it dynamically selected during runtime.

The IMS could be capable to understand and map O-Cloud ResourceTypes and their used ICS parameters that control their interconnections i.e., their network ports whenever the O-Cloud supports networking. This information could be used to automate or ensure that the interconnected resources are functional together when they are used as Assigned Resources e.g., in Node Clusters. Automation could be done e.g., through IMS selection of networking ResourceType ICS for the PortGroups. An example of matching ResourceType ICS is a server resource that need to have its NIC ports setup in a speed and termination type that matches the set of ports it connects to on the connected switch resources ports as exemplified in the figure below.

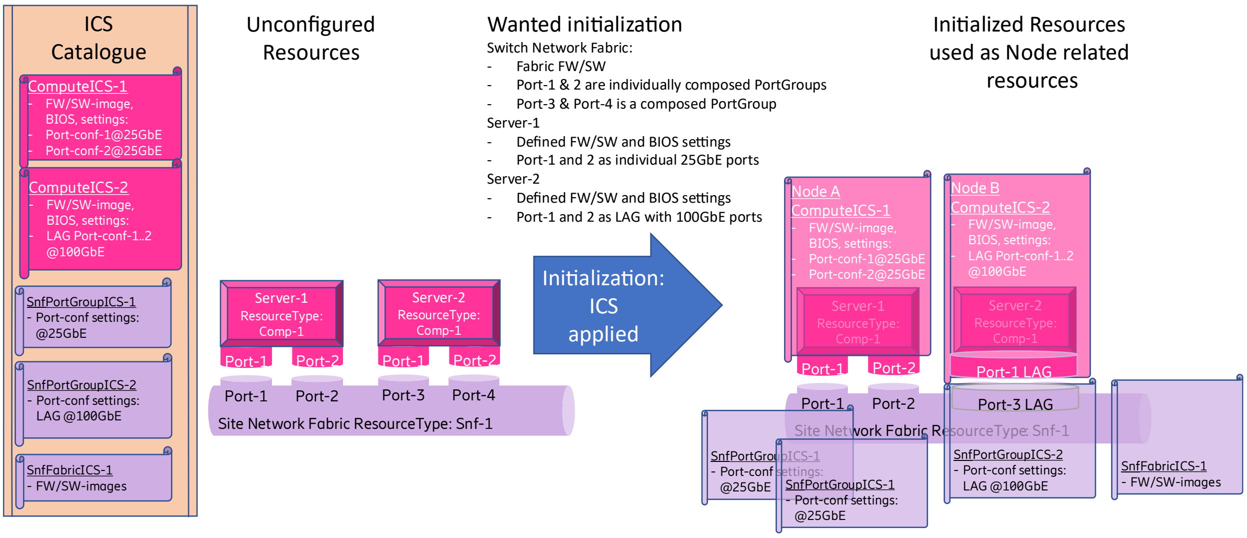

ICS Catalogue
ComputeICS-1
- FW/SW-image, BIOS, settings:
- Port-conf-1@25GbE
- Port-conf-2@25GbE
ComputeICS-2
- FW/SW-image, BIOS, settings:
- LAG Port-conf-1..2 @100GbE
SnfPortGroupICS-1
- Port-conf settings: @25GbE
SnfPortGroupICS-2
- Port-conf settings: LAG @100GbE
SnfFabricICS-1
- FW/SW-images
Unconfigured Resources
Server-1 ResourceType: Comp-1
Server-2 ResourceType: Comp-1
Port-1
Port-2
Port-1
Port-2
Port-1
Port-2
Port-3
Port-4
Site Network Fabric ResourceType: Snf-1
Wanted initialization
Switch Network Fabric:
- Fabric FW/SW
- Port-1 & 2 are individually composed PortGroups
- Port-3 & Port-4 is a composed PortGroup
Server-1
- Defined FW/SW and BIOS settings
- Port-1 and 2 as individual 25GbE ports
Server-2
- Defined FW/SW and BIOS settings
- Port-1 and 2 as LAG with 100GbE ports
Initialization: ICS applied
Initialized Resources used as Node related resources
Node A
ComputeICS-1
- FW/SW-image, BIOS, settings:
- Port-conf-1@25GbE
- Port-conf-2@25GbE
Server-1 ResourceType: Comp-1
Node B
ComputeICS-2
- FW/SW-image, BIOS, settings:
- LAG Port-conf-1..2 @100GbE
Server-2 ResourceType: Comp-1
Port-1
Port-2
Port-1 LAG
Port-1
Port-2
Port-3 LAG
Site Network Fabric ResourceType: Snf-1
SnfPortGroupICS-1
- Port-conf settings: @25GbE
SnfPortGroupICS-1
- Port-conf settings: @25GbE
SnfPortGroupICS-2
- Port-conf settings: LAG @100GbE
SnfFabricICS-1
- FW/SW-images

Figure C-1 Example of ICS Concept Usage

Annex D (normative): O-Cloud capabilities identification for use in LCM for the NF Deployment

The O-Cloud and O-Cloud Node Clusters, in particular, have different capabilities, resources with different capabilities, and a subset of these capabilities are relevant to be understood by the SMO for use in homing and the lifecycle management of the NF Deployments.

The standard O-Cloud capabilities identification aligns to what K8s resources can carry as part of the manifest (labels, annotations).

These O-Cloud capabilities include, non-exhaustively:

- AAL Profiles available (“aalProfileName” as defined in O-RAN.WG6.AAL Common API [x]),
- The labels that might be defined by the Node Feature Discovery (https://github.com/kubernetes-sigs/node-feature-discovery), feature labels which are referenced in the ASD TS [22]. Note that this does not include/reference the NFD daemon/mechanism as an implementation for provisioning these labels,
- Additional capabilities for the O-Cloud Node Clusters to support O-RAN features necessary for NF Deployments (e.g., the installed CNIs which extend the infrastructure capabilities such as secondary container cluster networking, etc).

Annex (informative):
Change history/Change request (history)

Date	Revision	Description
2023.11.13	05.00.01	Implemented CRs ERI-0001-Networking, ERI-0021, NOK-0074, NOK-0050, and editorial updates
2023.11.17	05.00.02	Editorial updates based on review comments
2023.11.28	06.00	Final version 06.00
2024.03.19	06.00.01	Implemented CRs DTAG-0002, DTAG-0003, and editorial updates
2024.03.22	06.00.02	Editorial updates based on review comments
2024.04.01	07.00	Final version 07.00
2024.07.15	07.00.01	Implemented CRs ERI-0035, ERI-0033, ERI-0031, DTAG-004, DTAG-006, DCM-0006
2024.07.18	07.00.02	Editorial updates based on review comments
2024.07.29	08.00	Final version 08.00
2024.11.14	08.00.01	Implemented CR DCM-0006, and editorial updates
2024.11.26	08.00.02	Editorial updates based on review comments
2024.12.09	08.01	Final version 08.01
2025.07.11	08.01.01	Implemented CRs ERI-0076, DCM-0007, ERI-0053, ERI-0073, ERI-0077, NOK-0159, NOK-0181
2025.07.17	08.01.02	Editorial updates based on review comments
2025.07.18	09.00	Final version 09.00