CNTT Edge Architecture -- Please see CNTT RM PR #2118

~~(RM Chapter 3 new section on Edge Computing w/o OpenStack specifics)~~

Alt: **RM Chapter08 – PR #2118** | Please make your comments on the PR and not here.

X.1 Introduction

Chapter 3(https://github.com/cntt-n/CNTT/blob/master/doc/ref_model/chapters/chapter01.md) of this document focuses on cloud infrastructure abstractions. While these are generic abstractions they and the associated capabilities are specified for data center or a colocation center cloud infrastructure. The environmental conditions, facility and other constraints, and the variability of deployments on the edge are significantly different and, thus, requires separate consideration.

It is unrealistic to expect that a private cloud can cost effectively meet the need of all loads, including peak and disaster recovery. It is for that reason that enterprises will implement an hybrid cloud. In a hybrid cloud deployment, at least two or more distinct cloud infrastructures are inter-connected together. In a multi-cloud the distinct cloud infrastructures of the hybrid cloud may be implemented using one or more technologies. The hybrid multi-cloud infrastructure has differences requiring different abstractions. These hybrid multi-clouds can be considered to be federated.

In IaaS clouds, the cloud infrastructure is defined but the tenant workloads include certain needed services (such as LB, messaging); thus, the VNF/CNFs may incorporate different services with the resultant issues related to an explosion of services, their integration and management complexities. To mitigate these issues, the CNTT Reference Model must specify the common services that every Telco cloud must support and thereby require workload developers to utilise these pre-specified services.

A generic Telco cloud is an hybrid multi-cloud or a Federated cloud that has deployments in large data centers, central offices or colocation facilities, and the edge. In this chapter we will discuss the characteristics of Telco Edge and hybrid multi-cloud.

X.2 Telco Edge Cloud

This section presents the characteristics and capabilities of different Edge cloud deployment locations, infrastructure, footprint, etc.

X.2.1 Telco Edge Cloud Deployment Environment Characteristics

Telco Edge Cloud (TEC) deployment locations can be **environmentally friendly** such as indoors (offices, buildings, etc.) or **environmentally challenged** such as outdoors (near network radios, curbside, etc.) or environmentally harsh environments (factories, noise, chemical, heat and electromagnetic exposure, etc). Some of the more salient characteristics are captured in Table X.1.

Table X.1. TEC Deployment Location Characteristics & Capabilities

	Facility Type	Environmental Characteristics	Capabilities	Physical Security	Implications	Deployment Locations	Comments
Environmentally friendly	Indoors: typical commercial or residential structures	Protected Safe for common infrastructure	Easy access to continuous electric power High/Medium bandwidth Fixed and /or wireless network access	Controlled Access	Commodotised infrastructure with no or minimal need for hardening /ruggedisation Operational benefits for installation and maintenance	Indoor venues: homes, shops, offices, stationary and secure cabinets Data centers, central offices, co-location facilities, Vendor premises, Customer premises	
Environmentally challenged	Outdoors and/or exposed to environmentally harsh conditions	maybe unprotected Exposure to abnormal levels of noise, vibration, heat, chemical, electromagnetic pollution	May only have battery power Low/Medium bandwidth Fixed and /or mobile network access	No or minimal access control	Expensive ruggedisation Operationally complex	Example locations: curbside, near cellular radios,	

X.2.2 Telco Edge Cloud Infrastructure Characteristics

Commodity hardware is only suited for environmentally friendly environments. Commodity hardware have standardised designs and form factors. Cloud deployments in data centers typically use such commodity hardware with standardised configurations resulting in operational benefits for procurement, installation and ongoing operations.

In addition to the type of infrastructure hosted in data center clouds, facilities with smaller sized infrastructure deployments, such as central offices or co-location facilities, may also host non-standard hardware designs including specialised components. The introduction of specialised hardware and custom configurations increases the cloud operations and management complexity.

At the edge, the infrastructure may further include ruggedised hardware for harsh environments and hardware with different form factors.

X.2.3 Telco Edge Cloud Infrastructure Profiles

The Reference Model (https://github.com/cntt-n/CNTT/blob/master/doc/ref_model/chapters/chapter04.md#4.2.4) specifies two infrastructure profiles:

> The Basic cloud infrastructure profile is intended for use by both IT and Network Function workloads that have low to medium network throughput requirements.
>
> The Network Intensive cloud infrastructure profile is intended for use by applications that have high network throughput requirements (up to 50Gbps).

The Network Intensive profile can specify extensions for hardware offloading; please see Hardware Acceleration Abstraction (https://github.com/cntt-n/CNTT/blob/master/doc/ref_model/chapters/chapter03.md#3.8). The Reference Model Network Intensive profile includes an initial set of Network Intensive profile extensions (https://github.com/cntt-n/CNTT/blob/master/doc/ref_model/chapters/chapter04.md#42421-network-acceleration-extensions).

Based on the infrastructure deployed at the edge, the Table X.2 specifies the Infrastructure Profile features and requirements (https://github.com/cntt-n/CNTT/blob/master/doc/ref_model/chapters/chapter05.md) that would need to be relaxed.

Table X.2. Characteristics of Infrastructure nodes

Reference	Feature	Description	As Specified in RM Chapter 05		Exception for Edge	
			Basic Type	Network Intensive	Basic Type	Network Intensive
infra.stg.cfg.003	Storage with replication		N	Y	N	Optional
infra.stg.cfg.004	Storage with encryption		Y	Y	N	Optional
infra.hw.cpu.cfg.001	Minimum Number of CPU sockets	This determines the minimum number of CPU sockets within each host	2	2	1	1
infra.hw.cpu.cfg.002	Minimum Number of cores per CPU	This determines the number of cores needed per CPU.	20	20	2	2
infra.hw.cpu.cfg.003	NUMA alignment	NUMA alignment support and BIOS configured to enable NUMA	N	Y	N	Optional

X.2.3 Telco Edge Cloud Infrastructure Characteristics

This section characterises the hardware capabilities for different edge deployments and the Platform services that run on the infrastructure. Please note, that the Platform services are containerised to save resources, and benefit from intrinsic availability and auto-scaling capabilities.

Table X.3. Characteristics of Infrastructure nodes

	Platform Services							Storage			Network Interfaces		
	Identity	Image	Placement	Compute	Networking	Message Queue	DB Server	Ephemeral	Persistent Block	Persistent Object	Management	Underlay (Provider)	Ove
Control Nodes	✓	✓	✓	✓	✓	✓	✓		✓		✓	✓	✓
Workload Nodes (Compute)				✓	✓			✓	✓	✓	✓	✓	✓
Storage Nodes									✓	✓	✓	✓	✓

Depending on the facility capabilities, deployments at the edge may be similar to one of the following:

> Small footprint edge device
>
> Single server: deploy multiple (one or more) workloads
>
> Single server: single Controller and multiple (one or more) workloads
>
> HA at edge (at least 2 edge servers): Multiple Controller and multiple workloads

X.2.4. Comparison of Edge terms from various Open Source Efforts

	Characteristics		Othe

CNTT Term?	Compute	Storage	Networking	RTT*	Security	Scalability	Elasticity	Resiliency	Preferred Workload Architecture	Upgrades		OpenStack	OPNFV Edge
Regional Data Center (DC) Fixed	1000's Standardised >1 CPU >20 cores /CPU	10's EB Standardised HDD and NVMe Permanence	>100 Gbps Standardised	~100 ms	Highly Secure	Horizontal and unlimited scaling	Rapid spin up and down	Infrastructure architected for resiliency Redundancy for FT and HA	Microservices based Stateless Hosted on Containers	HW Refresh: ? Firmware: When required Platform SW: CD		Central Data Center	
Metro Data Centers Fixed	10's to 100's Standardised >1 CPU >20 cores /CPU	100's PB Standardised NVMe on PCIe Permanence	> 100 Gbps Standardised	~10 ms	Highly Secure	Horizontal but limited scaling	Rapid spin up and down	Infrastructure architected for some level of resiliency Redundancy for limited FT and HA	Microservices based Stateless Hosted on Containers	HW Refresh: ? Firmware: When required Platform SW: CD		Edge Site	Large Edge
Edge Fixed / Mobile	10's Some Variability >=1 CPU >10 cores /CPU	100 TB Standardised NVMe on PCIe Permanence / Ephemeral	50 Gbps Standardised	~5 ms	Low Level of Trust	Horizontal but highly constrained scaling, if any	Rapid spin up (when possible) and down	Applications designed for resiliency against infra failures No or highly limited redund ancy	Microservices based Stateless Hosted on Containers	HW Refresh: ? Firmware: When required Platform SW: CD		Far Edge Site	Medium Edge
Mini- /Micro- Edge Mobile / Fixed	1's High Variability Harsh Environments 1 CPU >2 cores/CPU	10's GB NVMe Ephemeral Caching	10 Gbps Connectivity not Guaranteed	<2 ms Located in network proximity of EUD /IoT	Untrusted	Limited Vertical Scaling (resizing)	Constrained	Applications designed for resiliency against infra failures No or highly limited redund ancy	Microservices based or monolithic Stateless or Stateful Hosted on Containers or VMs Subject to QoS, adaptive to resource availability, viz. reduce resource consumption as they saturate	HW Refresh: ? Firmware: ? Platform SW: ?		Fog Computing (Mostly deprecated terminology) Extreme Edge Far Edge	Small Edge

```
*RTT: Round Trip Times
 EUD: End User Devices
 IoT: Internet of Things
```

++

+ OLD CONTENT +

++

Edge deployment scenarios

Cloud Infrastructure (CI) deployment environment for different edge deployments:

Controlled: Indoors, Protected, and Restricted environments. Data Centers, Central Offices, Indoor venues. Operational benefits for installation and maintenance, and reduced need for hardening/ruggedised.

Exposed: Outdoors, Exposed, Harsh and Unprotected environments. Expensive rugged equipment

Cloud Infrastructure (CI) hardware type for different edge deployments:

Commodity/Standard: COTS, standard hardware designs and form factors. Deployed only in Controlled environments. Reduced operational complexity.

Custom/Specialised: non-standard hardware designs including specialised components, ruggedised for harsh environments and different form factors. Deployed in Controlled and/or Exposed environments. Operationally complex environment.

Cloud Infrastructure (CI) hardware specifications for different edge deployments:

CNTT Basic: General Purpose CPU; Standard/Commoditised Design.

CNTT Network Intensive: CNTT Basic + high speed user plane (low latency, high throughput); Standard//Commoditised Design.

CNTT Network Intensive+ : CNTT Network Intensive + optional hardware acceleration (compared with software acceleration can result in lower power use and smaller physical size); possible Custom Design (Please see HW Acceleration Abstraction (url?).

CNTT Network Intensive++ : CNTT Network Intensive + required hardware acceleration; Custom Design.

Server capabilities for different edge deployments and the ~~OpenStack~~ Platform services that run on these servers; the ~~OpenStack~~ Platform services are containerised to save resources, intrinsic availability and autoscaling:

Control nodes host the ~~OpenStack~~ Platform control plane components ~~(subset of Cloud Controller Services)~~, and needs certain capabilities:

~~OpenStack~~ Platform services: Identity ~~(keystone)~~, Image ~~(glance)~~, Placement, Compute ~~(nova)~~, Networking ~~(neutron)~~ ~~with ML2 plug-in~~

Message Queue, Database server

Network Interfaces: management, provider and overlay

~~Compute~~ Workload nodes host ~~a subset of the Compute Node Services~~:

~~Hypervisor~~ Virtualisation Services

~~OpenStack~~ Compute ~~nova-compute (~~creating/deleting instances~~)~~

~~OpenStack~~ Networking ~~neutron-l2-~~agents/interfaces, VXLAN, metadata agent, and any dependencies

Network Interfaces: management, provider and overlay

Local Ephemeral Storage

Storage Nodes host the cinder-volume service. Storage nodes are optional and required only for some specific Edge deployments that need large persistent storage:

Block storage ~~cinder-volume~~

Storage devices specific ~~cinder-volume~~ drivers

Cloud partitioning: ~~Host Aggregates,~~ Availability Zones

~~OpenStack Edge Reference Architecture provides more depth and details~~

Edge Deployments:

Small footprint edge device: only networking agents

Single server: deploy multiple (one or more) ~~Compute~~ Workload nodes

Single server: single Controller and multiple (one or more) ~~Compute~~ Workload nodes

HA at edge (at least 2 edge servers): Multiple Controller and multiple ~~Compute~~ Workload nodes

Network Access: fixed and/or wireless (5G/LTE, WiFi, etc.)

Deployment Locations:

On Premises

Colocation facility

Vendor premises

Customer Premises

External (curb-side, proximity to radio site, etc.)

SDN Networking support on Edge

Comparison of Edge terms from various Open Source Efforts

	Characteristics		Othe

CNTT Term?	Compute	Storage	Networking	RTT*	Security	Scalability	Elasticity	Resiliency	Preferred Workload Architecture	Upgrades		OpenStack	OPNFV Edge
Regional Data Center (DC) Fixed	1000's Standardised >1 CPU >20 cores /CPU	10's EB Standardised HDD and NVMe Permanence	>100 Gbps Standardised	~100 ms	Highly Secure	Horizontal and unlimited scaling	Rapid spin up and down	Infrastructure architected for resiliency Redundancy for FT and HA	Microservices based Stateless Hosted on Containers	HW Refresh: ? Firmware: When required Platform SW: CD		Central Data Center	
Metro Data Centers Fixed	10's to 100's Standardised >1 CPU >20 cores /CPU	100's PB Standardised NVMe on PCIe Permanence	> 100 Gbps Standardised	~10 ms	Highly Secure	Horizontal but limited scaling	Rapid spin up and down	Infrastructure architected for some level of resiliency Redundancy for limited FT and HA	Microservices based Stateless Hosted on Containers	HW Refresh: ? Firmware: When required Platform SW: CD		Edge Site	Large Edge
Edge Fixed / Mobile	10's Some Variability >=1 CPU >10 cores /CPU	100 TB Standardised NVMe on PCIe Permanence / Ephemeral	50 Gbps Standardised	~5 ms	Low Level of Trust	Horizontal but highly constrained scaling, if any	Rapid spin up (when possible) and down	Applications designed for resiliency against infra failures No or highly limited redundancy	Microservices based Stateless Hosted on Containers	HW Refresh: ? Firmware: When required Platform SW: CD		Far Edge Site	Medium Edge
Mini-/Micro-Edge Mobile / Fixed	1's High Variability Harsh Environments 1 CPU >2 cores/CPU	10's GB NVMe Ephemeral Caching	10 Gbps Connectivity not Guaranteed	<2 ms Located in network proximity of EUD /IoT	Untrusted	Limited Vertical Scaling (resizing)	Constrained	Applications designed for resiliency against infra failures No or highly limited redundancy	Microservices based or monolithic Stateless or Stateful Hosted on Containers or VMs Subject to QoS, adaptive to resource availability, viz. reduce resource consumption as they saturate	HW Refresh: ? Firmware: ? Platform SW: ?		Fog Computing (Mostly deprecated terminology) Extreme Edge Far Edge	Small Edge

*RTT: Round Trip Times
EUD: End User Devices
IoT: Internet of Things

Glossary

- State of the Edge