ETSI GS NFV-PER 001 V1.1.1 (2014-06)

Network Functions Virtualisation (NFV);
NFV Performance & Portability Best Practises

Disclaimer

This document has been produced and approved by the Network Functions Virtualisation (NFV) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference
DGS/NFV-PER001

Keywords
performance, portability

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the only prevailing document is the print of the Portable Document Format (PDF) version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://ipr.etsi.org).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Group Specification (GS) has been produced by ETSI Industry Specification Group (ISG) Network Functions Virtualisation (NFV).

Modal verbs terminology

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

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

1 Scope

The present document provides a list of features which the performance and portability templates (Virtual Machine Descriptor and Compute Host Descriptor) should contain for the appropriate deployment of Virtual Machines over a Compute Host (i.e. a "telco datacentre").

In addition, the document provides a set of recommendations and best practises on the minimum requirements that the HW and hypervisor should have for a "telco datacentre" suitable for data-plane workloads. The recommendations and best practises are based on tests results from the performance evaluation of data-plane workloads. It is recognized that the recommendations are required for VNFs supporting data plane workloads and that a small portion of the recommended list are not required in all cases of VNFs, such as VNFs related to control plane workloads.

2 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.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

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

2.1 Normative references

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

[1] ETSI GS NFV 003: "Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV".

[2] ETSI GS NFV 001: "Network Functions Virtualisation (NFV); Use Cases".

2.2 Informative references

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] Open Virtualisation Format Specification Version 2.1.0.

NOTE: Available at: http://www.dmtf.org/sites/default/files/standards/documents/DSP0243_2.1.0.pdf.

[i.2] Libvirt - The Virtualisation API.

NOTE: Available at: http://libvirt.org/.

[i.3] AWS® CloudFormation.

NOTE 1: Available at: http://aws.amazon.com/cloudformation/.

NOTE 2: AWS® CloudFormation is an example of a suitable product available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of this product.

[i.4] Portable Hardware Locality.

NOTE: Available at: http://www.open-mpi.org/projects/hwloc/.

[i.5] IETF RFC 2544: "Benchmarking Methodology for Network Interconnect Devices".

NOTE: Available at: http://www.ietf.org/rfc/rfc2544.txt.

[i.6] IETF RFC 2889: "Benchmarking Methodology for LAN Switching Devices".

NOTE: Available at: http://www.ietf.org/rfc/rfc2889.txt.

[i.7] IETF RFC 3918: "Methodology for IP Multicast Benchmarking".

NOTE: Available at: http://www.ietf.org/rfc/rfc3918.txt.

[i.8] IETF RFC 3511: "Benchmarking Methodology for Firewall Performance".

NOTE: Available at: http://tools.ietf.org/rfc/rfc3511.txt.

[i.9] IEEE 1588: "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document the terms and definitions given in GS NFV 003 [1] and the following apply:

NOTE: A term defined in the present document takes precedence over the definition of the same term, if any, in GS NFV 003 [1].

Network Function (NF): A functional building block within a network infrastructure, which has well-defined external interfaces and a well-defined functional behaviour. In practical terms, a Network Function is today often a network node or physical appliance.

Network Functions Virtualisation Infrastructure (NFVI): The NFV-Infrastructure is the totality of all hardware and software components which build up the environment in which VNFs are deployed. The NFV-Infrastructure can span across several locations, i.e. multiple N-PoPs. The network providing connectivity between these locations is regarded to be part of the NFV-Infrastructure. The NFV-Infrastructure includes resources for computation, networking and storage.

Compute Host: A Compute Host is the whole server entity, part of an NFVI, composed of a HW platform (processor, memory, I/O devices, internal disk) and a hypervisor running on it.

Compute Host Descriptor (CHD): A Compute Host Descriptor is a template to define a storage schema for the capabilities and up-to-date available resources which can be offered by a Compute Host server to VM Images at deployment time. Therefore, there will be one Compute Host Descriptor for every Compute Host, containing both its capabilities and its available resources.

Virtualised Network Function (VNF): An implementation of an NF that can be deployed on a Network Function Virtualisation Infrastructure (NFVI). A VNF can be deployed as a set of Virtual Machines (VM), as SW components deployable, maintainable and manageable, emulating a single computer.

Virtual Machine Instance (VM Instance): A Virtual Machine Instance is a Virtual Machine already running in a Compute Host.

Virtual Machine Image (VM Image): A VM Image is an executable SW component, subject to be deployed in Compute Hosts as one or several Virtual Machine Instances.

Virtual Machine Configuration (VM Configuration): A VM Configuration is the final configuration to be applied when deploying a VM Image on a specific Compute Host.

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

AAA	Authentication, Authorization, and Accounting
API	Application Programming Interface
ARP	Address Resolution Protocol
AS	Application Server
BBU	Base Band Unit
BFD	Bidirectional Forwarding Detection
BGP	Border Gateway Protocol
BIOS	Basic Input/Output System
BNG	Broadband Network Gateway
BRAS	Broadband Remote Access Server
BW	Bandwidth
CDN	Content Delivery Network
CGNAT	Carrier Grade Network Address Translation
CHD	Compute Host Descriptor
CIFS	Common Internet File System
COTS	Commercial Off-The-Shelf
CPE	Customer Premises Equipment
CPU	Central Processing Unit
C-RAN	Cloud-Radio Access Network
CVLAN	Customer VLAN
DCB	Data Center Bridging
DDoS	Distributed Denial of Service
DDR2	Double Data Rate type 2
DDR3	Double Data Rate type 3
DHCP	Dynamic Host Configuration Protocol
DMA	Direct Memory Access
DPI	Deep Packet Inspection
DSLAM	Digital Subscriber Line Access Multiplexer
DUT	Device Under Test
E-CPE	Enterprise-Customer Premises Equipment
ERPS	Ethernet Ring Protection Switching
FFT	Fast Fourier Transform
FIB	Forwarding Information Base
FTP	File Transfer Protocol
FW	Firewall
GB	GigaByte
GE	Gigabit Ethernet
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GPS	Global Positioning System
GRE	Generic Routing Encapsulation
GUI	Graphical User Interface
GW	Gateway
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
HW	Hardware
I/O	Input/Output
I-CSCF	Interrogating-Call Session Control Function
IMS	IP Multimedia Subsystem
IO	Input Output
IOMMU	Input/Output Memory Management Unit
IOTLB	I/O Translation Lookaside Buffer
IP	Internet Protocol
IPC	Inter-Process Communication
IPoE	IP over Ethernet
IPsec	IP security
ISIS	Intermediate System to Intermediate System

KPI	Key Performance Indicator
L4	Layer 4
L7	Layer 7
LPM	Longest Prefix Match
MAC	Media Access Control
MAN	Metropolitan Area Network
MANO	MANagement and Orchestration
MGCF	Media Gateway Controller Function
MME	Mobility Management Entity
MMU	Memory Management Unit
MOS	Mean Opinion Score
MOS-AV	MOS-Audio & Video
MPLS	Multi-Protocol Label Switching
MSE	Mean Square Error
N/A	Not Applicable
NAS	Network-Attached Storage
NAT	Network Address Translation
NF	Network Function
NFV	Network Functions Virtualisation
NFVI	Network Functions Virtualisation Infrastructure
NIC	Network Interface Card
NUMA	Non-Uniform Memory Access
OLT	Optical Line Terminal
ONT	Optical Network Terminal
ONU	Optical Network Unit
OS	Operating System
OSPF	Open Shortest Path First
OVF	Open Virtualisation Format
P2P	Peer-to-Peer
PCI	Peripheral Component Interconnect
PCIe	PCI Express
PCIe	VF PCIe Virtual Function
PCI-SIG	PCI Special Interest Group
PCRF	Policy and Charging Rules Function
P-CSCF	Proxy-Call Session Control Function
PDN	Packet Data Network
PE	Provider Edge
P-GW	PDN-Gateway
PNF	Physical Network Function
PPP	Point-to-Point Protocol
PPPoE	Point-to-Point Protocol over Ethernet
PSNR	Peak Signal-to-Noise Ratio
QoE	Quality of Experience
QoS	Quality of Service
R/W	Read/Write
RADIUS	Remote Authentication Dial In User Service
RAM	Random Access Memory
RAN	Radio Access Network
RDMA	Remote Direct Memory Access
RGW	Residential Gateway
RIB	Routing Information Base
RMS	Root Mean of Squares
RoCE	RDMA over Converged Ethernet
RX	Reception
SAN	Storage Area Network
S-CSCF	Serving-Call Session Control Function
SGSN	Serving GPRS Support Node
S-GW	Serving-Gateway
SLA	Service Level Agreement
SMT	Simultaneous Multi-Threading
SR-IOV	Single Root I/O Virtualisation
SSIM	Structural Similarity

STB	Set-Top-Box
STP	Spanning Tree Protocol
SVLAN	Service VLAN
SW	Software
TCP	Transmission Control Protocol
TLB	Translation Lookaside Buffer
TWAMP	Two-Way Active Measurement Protocol
TX	Transmission
VF	Virtual Function
VIA	Virtual Interface Architecture
VIM	Virtualised Infrastructure Manager
VLAN	Virtual Local Area Network
VM	Virtual Machine
VNF	Virtualised Network Function
VQM	Video Quality Metric
VTA	Virtual Test Appliance
EMS	Element Management System
KVM	Kernel-based Virtual Machine
DNS	Domain Name System
NTP	Network Time Protocol
SSH	Secure SHell
NFS	Network File System
TA	Test Agent
ISV	Independent Software Vendor
ABR	Available Bit Rate
CBR	Constant Bit Rate
VBR	Variable Bit Rate
RSS	Receive Side Scaling

4 Introduction

The present document provides a list of minimal features which the VM Descriptor and Compute Host Descriptor should contain for the appropriate deployment of VM Images over an NFVI (i.e. a "telco datacentre"), in order to guarantee high and predictable performance of data plane workloads while assuring their portability. In addition, the document provides a set of recommendations on the minimum requirements which HW and hypervisor should have for a "telco datacentre" suitable for different workloads (data-plane, control-plane, etc.) present in VNFs.

It should be noted that the present document focuses on capturing HW and SW requirements for VMs dealing with data-plane workloads.

As shown in figure 1, a VNF can be composed of a set of VMs , and their internal and external interfaces.

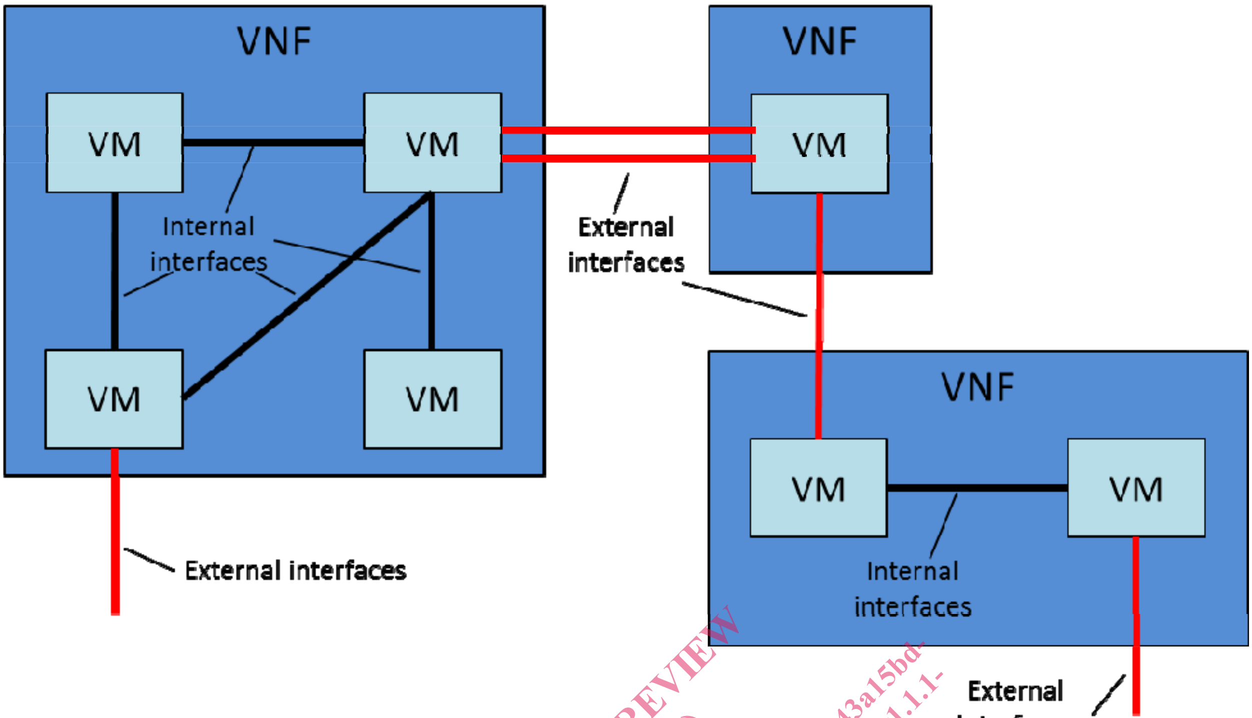

Figure 1: VNF and VMs

Regarding NFV-Infrastructure, the document will focus on NFVI compute domain, leaving storage and networking HW domains for further analysis. Regarding NFV SW, only hypervisor capabilities are considered at this stage. The term "**Compute Host**" will be used to refer to the whole entity composed of the HW platform (processor, memory, I/O devices, internal disk) and a hypervisor running on it. Scenarios with two or more hypervisors on the same server are not considered for this first analysis.

In the present document, the term Compute Host Descriptor (CHD) will be used - applied to every Compute Host - to describe the template defining a storage schema for the capabilities and currently available resources offered by the Compute Host for the deployment of VM Images. Hence, there will be one CHD for every Compute Host, containing both its capabilities and its available resources. In a practical realization, since the capabilities can be the same for a common type of Compute Host (HW and hypervisor), the CHD could be implemented as the combination of two different templates -one containing the capabilities, the other one containing the available resources-, thus being possible to have a common template for the capabilities of a given type of Compute Host. Along the document, the term CHD will sometimes be used to refer to one of its potential practical realizations: a parseable template file describing both the key HW capabilities and available resources of server Compute Host.

On the other hand, for the purposes of performance and portability discussions, it is required a term to describe the NFV resources which are required by a VM Image from the NFVI. Additionally, for a proper discussion on the subject, it is required a term to distinguish the demanded resources from the applied configuration. For this purpose, in the present document the following new terms are coined:

- **Virtual Machine Descriptor** (VM Descriptor) is a template which declares the NFV resources to be required by the VNF's VMs from the NFVI. It is a type of meta-information associated with a VM Image. The description is made in the form of a SW template, which is computable by the MANO Functions. E.g. if OVF [i.1] is used as the way of packaging VM images, then VM Descriptor could be part of the OVF meta-data. It is expected a single VM Descriptor for each VM Image type, independently on the number of VM Instances deployed on an NFVI.

- **Virtual Machine Configuration** (VM Configuration) is the final configuration to be applied when deploying a VM Image on a specific Compute Host. E.g. If libvirt [i.2] was being used by the hypervisor manager, then the VM Configuration could be analogous to the *vm.xml* that is generated at deployment (containing host-specific information such as PCI device ids, core ids, etc.). It is expected a VM configuration for each deployed VM Instance.

Along the document, for the sake of simplicity, the term VM Descriptor might also be used as synonymous of one of its potential practical realizations: a parseable template file describing the requirements of a VM Image. Likewise, several VM Descriptors could be joined together in a "**template file**" (e.g. analogous to AWS® CloudFormation templates [i.3], see note) describing the whole VNF - comprising its components (VMs), the internal interfaces between VMs and the set of external interfaces.

NOTE: AWS® CloudFormation is an example of a suitable product available commercially. This information is given for the convenience of users of the present document and does not constitute an endorsement by ETSI of this product.

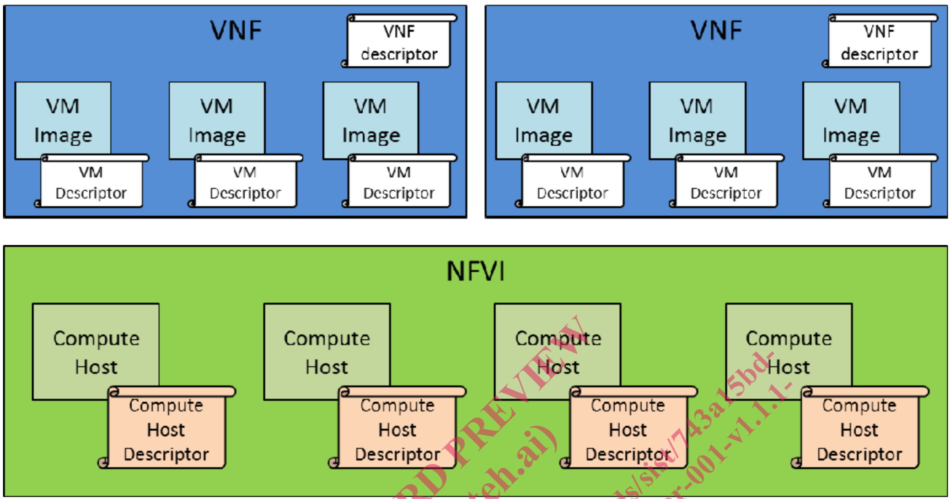

Figure 2: Compute Host Descriptor (CHD) and VM Descriptor templates

We will assume scenarios where VNFs might be provided by VNF SW vendors as SW packages including a descriptor for the whole VNF, a set of VM Images, and their corresponding VM Descriptors. In the simplest scenario of deployment of a VNF, with a VNF composed of a single VM Image, and after all the preparatory steps (e.g. inventory check), the MANO Functions would read the VM Descriptor and parse the requirements for the underlying HW. Next, it would check what Compute Hosts would be suitable to deploy the VM Image, based on the different Compute Host Descriptors. Those CHDs should show up-to-date availability of resources (free CPUs, free memory, etc.). From the set of valid Compute Hosts, the MANO Functions will choose one (the nature of this selection procedure is not relevant for the purpose of the present document), and, then, a VM Configuration will be elaborated, containing the specific resources that should be consumed and configured on the selected Compute Host. That VM Configuration will be used to deploy the VM Image on the selected Compute Host and, therefore, is specific for that server since it may ask for specific devices or resources. Finally, the VM Image is deployed on the server, launched by the MANO Functions.

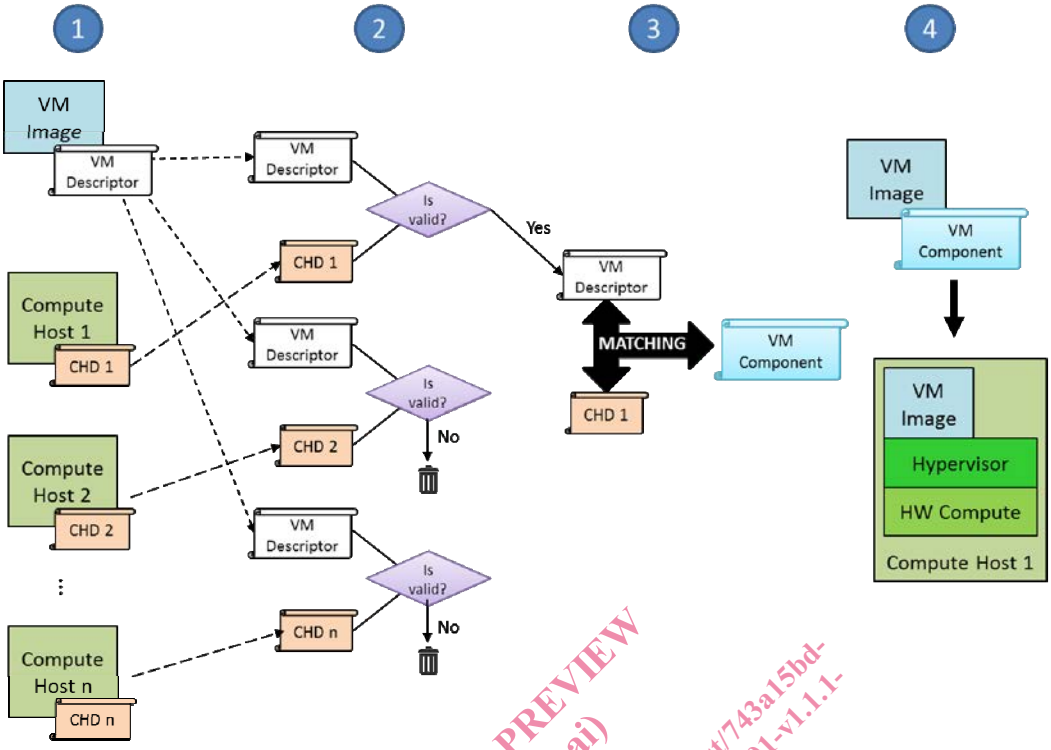

Figure 3: Deployment process as part of the orchestration process

The purpose of the present document is to provide the list of VM requirements that should be included in the VM Descriptor template, and the list of HW capabilities that should be included in the Compute Host Descriptor (CHD) to assure predictable high performance. It is assumed that the MANO Functions will make the mix & match. The details of the orchestration process, the VM Configuration and the specific format of the CHD and VM Descriptor templates are out of the scope of the present document.

The present document is structured as follows. First, clause 5 describes the methodology used to identify the list of requirements and capabilities for the VM Descriptor and CHD templates. The methodology is based on experimentation through tests and comparison of the test results in both bare metal and virtualisation environments in order to identify HW and SW bottlenecks. Instead of testing all the possible VNFs, a pragmatic approach has been followed, focusing on significant VNFs already described in the NFV use cases [2]. Moreover, instead of testing the whole VNF, tests have been restricted to specific tasks or workloads, which happen to be present in several VNFs. Relevant workloads in terms of performance have been identified and are also described in clause 5. The clause also introduces a first collection of tests (completed or planned) covering some of these workloads.

Next, clause 6 presents the main conclusions for each tested workload, highlighting those relevant HW and SW features and how they affect performance. The present document covers specifically the analysis of data plane workloads in intermediate elements.

Based on the workload analysis in clause 6, clause 7 provides a recommendation on the minimum requirements that a Compute Host should have for a "telco datacentre" suitable for data-plane workloads, and, finally, clause 8 gathers the list of features which the Compute Node Descriptor and the VM Descriptor templates should contain for the appropriate deployment of VM Images over an NFVI (i.e. a "telco datacentre"), in order to guarantee high and predictable performance while preserving portability across different servers.