BSP System Dimensioning Guide, CEE R6

Cloud Execution Environment

CONFIGURATION MODEL

2/192 02-AZE 102 01/5-V3 Uen A

Contents

1 Introduction

This document describes the characteristics of Cloud Execution Environment (CEE) to enable dimensioning and understanding the limitations of CEE. It also describes requirements on HW combinations required for running CEE. The application can have additional requirements.

Storage is measured in gibibyte (GiB), tebibyte (TiB), and mebibyte (MiB) in this document.
1 GiB is equivalent to 1.074 GB.

The following words are used in this document with the meaning specified below:

vNIC	A virtual network interface card (vNIC) provides connectivity between the Cloud SDN Switch (CSS) and a VM. A configuration can provide several vNICs to a VM.
Interface	A network interface. Can be either a physical NIC (PHY) providing CSS with board external connectivity, or a virtual NIC connecting CSS to a VM.
PMD thread	CSS uses a Poll Mode Driver (PMD) technique that continuously polls incoming packets from the NICs, that is, interrupts are not used. To be able to reliably handle all incoming packets, a software continuously polls the NIC queues for packets to be handled. This software is executing in one or more threads that are called PMD threads. The execution environment for the PMD threads is isolated from the Linux scheduler to be able to reliably handle a high sustained packet flow without interrupts or delays caused by being scheduled out.
vCIC host	A host running Compute and vCIC. There are three vCIC hosts in Multi-Server deployments of CEE.
Compute host	A host running Compute without vCIC

1.1 Target Group

Cloud Infrastructure providers and application designers.

1.2 System Characteristics

The characteristic features of CEE are the following:

- High Availability cloud system
- OpenStack® SW deployment (for deviations, see Section 5.1 on page 30)

For information on system limitations, see Section 5 on page 30.

For BSP characteristics refer to the BSP product documentation.

2 CEE System

This section describes CEE system configurations and capabilities.

2.1 System Configurations

CEE is a scalable system used with HW products from different vendors. CEE offers a set of configurations.

2.1.1 Reference Configurations

Reference configurations offer the possibility to build and scale a cloud system on specified HW, such as servers, switches, and storage.

BSP can include one to six subracks with a maximum of 12 blades in each subrack, therefore, the biggest available configuration can contain up to 72 blades altogether in six subracks.

2.1.2 Certified Configuration

Certified configurations are a subset of reference configurations. Certified CEE configurations have been documented and tested by the Ericsson Cloud organization.

The certified configuration for CEE R6 consists of the following HW components:

- BSP cabinet with 6 subracks

3 HW Requirements

This section describes generic hardware and firmware requirements for CEE, based on the certified Ericsson CEE R6 HW.

Complying with the outlined HW requirements does not in any way imply Ericsson CEE certification. HW deployment outside the already certified ones requires system integration work.

The server, switching, and optional storage components of the HW environment are shown in Figure 1:

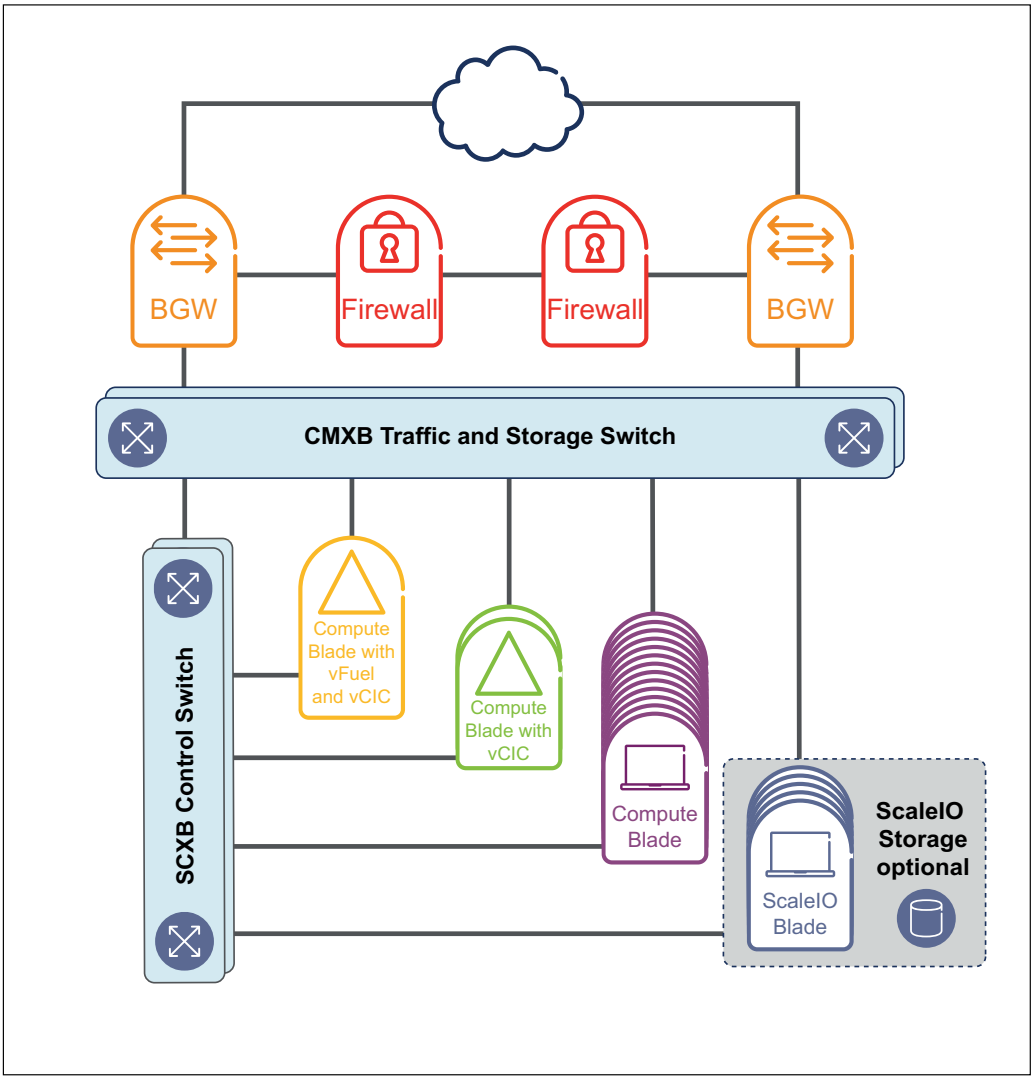

Figure 1 CEE HW Environment with BSP HW

For information on the supported BSP 8100 HW, see Table 41. For more information, refer to the BSP document: BSP Technical Product Description, Reference [1].

3.1 Summary of Blade Configuration

This section describes HW requirements for blades.

Table 1 HW Configuration Example, GEP5

Aspect	Requirement
CPU	1 x Intel Xeon E5-2658 v2 processor (10 cores per processor, 2 Hyperthreads per core), 20 Hyperthreads available
RAM	64 GiB
NIC	4*10GE Intel Niantic and 2*1GE unspecified (Management)
Onboard Disk	• 1200 GiB SSD (GEP5-64-1200) • 400 GiB DISK (GEP5-64-400) • 8 GiB (GEP5-64)
Management Interface	EBS/BSP Out of band management

Table 2 HW Configuration Example, GEP7

Aspect	Requirement
CPU	1 x Intel Xeon E5-2658 v4 processor (14 cores per processor, 2 Hyperthreads per core), 28 Hyperthreads available
RAM	128 GiB
NIC	4*10GE Fortville
Onboard Disk	• 1600 GiB SSD (GEP7-128-X16) • 8 GiB (GEP7-128-X
Management Interface	EBS/BSP Out of band management

Note: SR-IOV is not supported for BSP HW.

For the characteristics of the GEP blade refer to the BSP documentation.

3.2 Network Configuration

This section describes HW configuration for networking.

For network configuration refer to the BSP documentation.

3.2.1 Network Requirements for Distributed Storage, ScaleIO

The following requirements must be fulfilled:

Network:

- Bandwidth:
  - Minimal: 2 x 10 Gbit
  - Optimal: 4 x 10 Gbit, to separate front-end and back-end traffic
  - Additional 2 x 1 Gbit network for management traffic
- All the ScaleIO components are accessible via the 10 Gbit network.
- Network bandwidth and latency between all nodes is acceptable, according to application demands.
- The Ethernet switch supports the required bandwidth between network nodes.
- MTU settings are consistent across all servers and switches.
- The following TCP ports are not used by any other application, and are open in the local firewall of the server:
  - Meta Data Manager (MDM): 6611 and 9011
  - ScaleIO Data Server (SDS): 7072, for multiple SDS: 7073-7076
  - ScaleIO Gateway (including REST Gateway, Installation Manager, and SNMP trap sender): 80 and 443
  - Light Installation Agent (LIA): 9099

Parameters and configuration in ScaleIO that will affect dimensioning:

- SDS network limits can be set to avoid overloading the storage network with huge amount of rebuild or rebalance traffic. Refer to the **SDS network limits** section in the *EMC ScaleIO Version 2.0.x User Guide*.

3.3 Firewall Requirement

During test of the CEE Certified Configuration, a Juniper EX 4550 has been used as BGW and a Juniper SRX 3400 as Firewall.

All Juniper SRX High-End series Firewalls provide a similar set of features. SRX can be equipped with a flexible number of Services Processing Cards (SPC), I/O Cards (IOC), and Network Processing Cards (NPC), allowing the system configuration to support a balance of performance and port density. Suitable firewall HW setup depends on expected traffic volumes and sessions. More information can be found in supplier datasheets, on the Juniper Networks homepage, Reference [2].

3.4 CPU Configuration

Refer to the *Configuration File Guide* for more information about the configuration procedure.

The number of available CPU IDs depends on the CPU model.

3.4.1 Compute Host

Table 3 and Figure 2 give an example for automatic CPU allocation for VMs, OVS, vCIC, and vFuel. This allocation is valid for GEP5 with Intel Xeon E5-2658 v2 processor.

If the number of cores in the used processor type differs from the number of cores in this example, the number of cores allocated to the VMs must be adjusted.

The example describes the CPU allocation at a server that runs vCIC and vFuel. If vCIC or vFuel are not used on the server, their CPUs are allocated for the tenant VMs.

Table 3 Automatic CPU Allocation on GEP5 with Intel Xeon E5-2658 v2 Processor, for VMs, OVS, vCIC, and vFuel in a Compute Host

CPU Owner	Allocated CPU ID
Tenant VM	8,18, 9,19
OVS(1)	1,11
OVS control process(2)	0(2)
vCIC(3)	4,14, 5,15, 6,16 7,17
vFuel	3,13

CPU Owner	Allocated CPU ID
Host OS(4)	0,10, 2,12
If ScaleIO is installed, then 1–5 CPU threads are used for SDC.	

(1) To achieve a more predictable performance, allocate only one CPU per core for OVS. See Section 3.4.1.1 on page 9 for more information.
(2) The process does not get a CPU for its exclusive use. A configuration parameter specifies one of the host OS CPUs to be used by the OVS control process.
(3) The vCIC can be allocated on cores that are over allocated and shared with the application. In such cases, it must be ensured that the application sharing resources with the vCIC dont exhaust the vCIC resources.
(4) In some use-cases only one core is allocated to the host OS as explained in Section 3.4.1.2 on page 9.

Table 4 Automatic CPU Allocation on GEP7 with Intel Xeon E5-2658 v4 Processor, for VMs, OVS, vCIC, and vFuel in a Compute Host

CPU Owner	Allocated CPU ID
Tenant VM	8,22, 9,23, 10,24, 11,25, 12,26, 13,27
OVS(1)	1,15
OVS control process(2)	0(2)
vCIC(3)	4,18, 5,19, 6,20 7,21
vFuel	3,17
Host OS(4)	0,14, 2,16
If ScaleIO is installed, then 1–5 CPU threads are used for SDC.	

(1) To achieve a more predictable performance, allocate only one CPU per core for OVS. See Section 3.4.1.1 on page 9 for more information.
(2) The process does not get a CPU for its exclusive use. A configuration parameter specifies one of the host OS CPUs to be used by the OVS control process.
(3) The vCIC can be allocated on cores that are over allocated and shared with the application. In such cases, it must be ensured that the application sharing resources with the vCIC dont exhaust the vCIC resources.
(4) In some use-cases only one core is allocated to the host OS as explained in Section 3.4.1.2 on page 9.

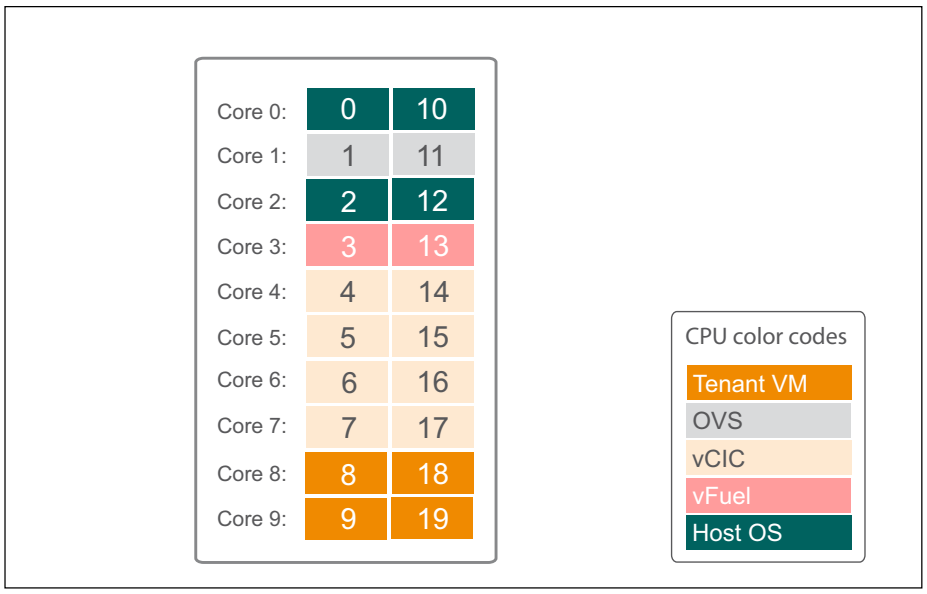

Figure 2 Automatic CPU Allocation of the Respective Resource Owner on GEP5 with Intel Xeon E5-2658 v2 Processor for VMs, OVS, vCIC, and vFuel in a Compute Host

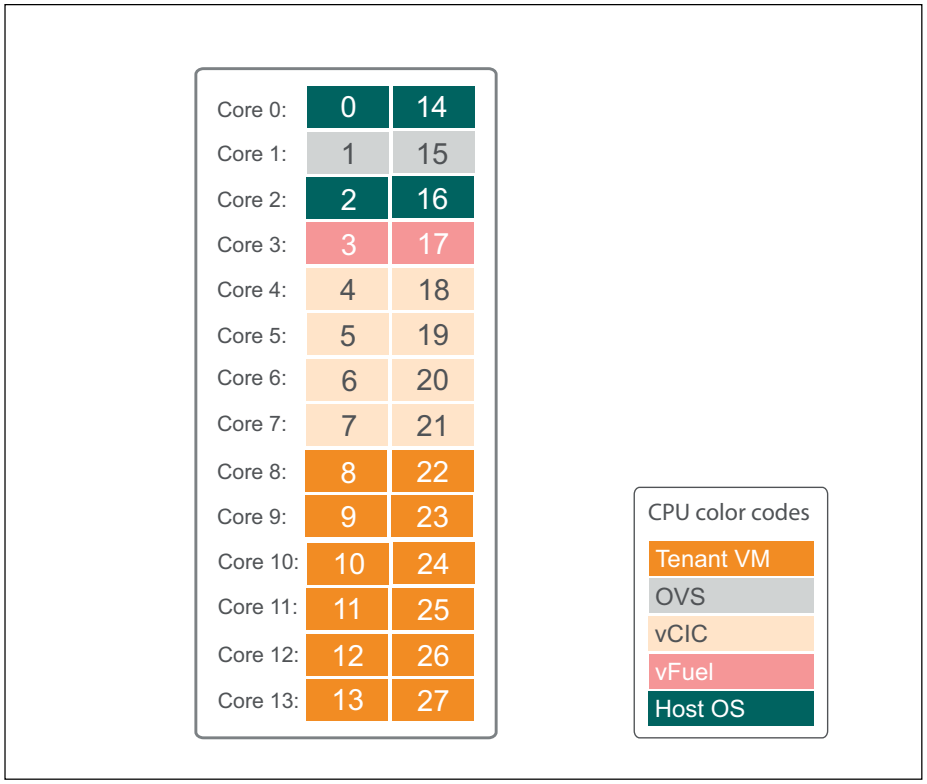

Figure 3 Automatic CPU Allocation of the Respective Resource Owner on GEP7 with Intel Xeon E5-2658 v4 Processor for VMs, OVS, vCIC, and vFuel in a Compute Host

3.4.1.1 Allocating Single CPUs for OVS

To achieve a more predictable OVS performance, only a single CPU must be allocated to OVS from each CPU core reserved for OVS PMD threads. The other CPU on the same core, also called hyper-thread sibling, must be isolated, that is, not used by the host OS process scheduler, and is free from any extra load that can negatively influence the OVS performance.

To achieve the described allocation, the following settings must be performed:

- Refer to the *Configuration File Guide* for the following setting in the configuration file:

  The other CPU in the core must be reserved for the `idle` owner. By this reservation, the following results are achieved:

  - The idle CPUs are listed in the kernel's `isolcpus` boot parameter. It ensures that the process scheduler does not assign any process to these CPUs.
  - The idle CPUs are listed in the kernel's `nohz_full` boot parameter. It ensures that the kernel does not generate scheduling clock interrupts on these CPUs.

- Refer to *SW Installation in Multi-Server Deployment* for the following setting at each Compute host:

  The idle CPUs must be listed in the `noirqs` parameter of the `configure-interrupts` upstart task in order to avoid these CPUs from using them for serving interrupt requests.

3.4.1.2 Allocating One Core to the Host OS

However the recommended CPU allocation to the host OS is two cores for each Compute host, for some use-cases, where a single or very few VMs are used on each Compute host, it can be possible to reduce the host OS allocation to one core.
Having a single core allocated for the host OS can impact the characteristics of the system. For example, actions such as starting, stopping, and migrating VMs can become slower, and it can also impact the performance of other VMs on the same host. Before using a system with a single core allocated to the Host OS in a production environment, sufficient testing must be performed. Beside other checks, this testing must include the capacity and performance test of the VM during the time of performing life cycle management operations for other VMs running on the same host. While performing such test, the processor load for the host OS must be monitored carefully for deviation from a steady state.

3.4.2 ScaleIO Host

Table 5 shows the minimum number of cores to be allocated to the CPU owners in a ScaleIO host.

Table 5 Minimum Number of Allocated Cores in a ScaleIO Host

CPU Owner	Minimum Number of Allocated Cores
Host OS	2
Meta Data Manager/Tie-Breaker (MDM/TB)	1
SDS	2–4

3.5 RAM Configuration

This section describes the optimal RAM configuration for the CEE.

Refer to the *Configuration File Guide* for more information about the configuration procedure.

Each Neutron network created consumes RAM in the vCIC, and it influences the maximum number of virtual tenant networks. See Section 4.4.3 on page 27 for more information.

The RAM size on GEP5-64-400 and GEP5-64-1200 boards is 64 GiB. The RAM size on GEP7-128-X16 and GEP7-128-X boards is 128 GiB.

Running vCIC, vFuel, or both on the server modifies the optimal memory allocation. The subsections below describe the possible cases for 128 and 64 GiB memory.

See the following cases:

- Server without vCIC and vFuel, see Section 3.5.2 on page 11.
- Server with vCIC, see Section 3.5.3 on page 13.
- Server with vFuel, see Section 3.5.4 on page 14.
- Server with vCIC and vFuel, see Section 3.5.5 on page 15.

Section 3.5.1 on page 10 provides general information about RAM configuration in CEE.

3.5.1 Introduction

The default configuration for the Host OS on Compute nodes is 8 GiB. More memory can be reserved for the Host OS by setting the relevant configuration parameters. The needed amount of memory depends on the values of a number of other parameters as described below.

The memory used for the VMs is allocated to huge pages. This is the memory visible from the inside of the VMs. The 1 GiB huge pages are referred to

as **Tenant VM** in the RAM reservation tables of the upcoming sections. In addition to the 1 GiB huge pages, the VMs need memory allocated from the host OS. This memory is used, for example, to emulate devices used by the virtual machine. It is hard to predict the amount of host OS memory used by the emulator since, for example, it depends on the type and number of the used devices. A small VM consumes less than 100 MiB, while it can grow to several hundred MiB in specific cases. About 300 MiB host OS memory would be enough for each virtual machine but we must double it and calculate with 600 MiB as explained below.

In a system using the NUMA architecture, the NUMA location of VMs must be considered. The available memory, that is, the huge pages and the Host OS memory, are evenly distributed between the NUMA nodes. By design, OpenStack Nova allocates VMs on the first NUMA node that fits the VM. Apart from the VMs running on both NUMA nodes, the VMs allocate memory from the NUMA node on which they are running. In a worst case scenario where all VMs are allocated on the same NUMA node, all the memory for the VMs will be allocated from the same NUMA node. In such a scenario most of the memory on the other NUMA node will be unused, and half of the memory on the Compute node will be free. To be on the safe side in a dual socket system, the 300 MiB host OS memory per VM must be doubled to cover the case where all VMs are allocated on the same NUMA node.

The host OS memory usage for processes other than the VMs depends on the CPU reservations. The host OS uses the unreserved CPUs. By default, CEE allocates two cores to the host OS on NUMA node 0. This means that most of the memory used by the Host OS will be allocated from NUMA node 0. In some scenarios it is preferred to run the host OS on both NUMA nodes. It can be achieved by modifying the CPU reservation.

Note: In order to allocate as little memory for the host OS as possible, memory profiling of the host OS for the specific scenario is recommended.

3.5.2 Compute Server without vCIC and vFuel

Table 6 and Table 7 specify the volumes allocated to the resource owners. Host Operating System (OS) is not included, so the tables contain 120 and 56 GiB, respectively.

Table 8 and Table 9 contain the total memory sizes.

Table 6 Memory Allocation for System with 128 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
Tenant VM(1)	1024	118	118
OVS	2	1024	2

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
vCIC	-	-	-
vFuel	-	-	-

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

Table 7 Memory Allocation for System with 64 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
Tenant VM(1)	1024	54	54
OVS	2	1024	2
vCIC	-	-	-
vFuel	-	-	-

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

Table 8 Total Memory Sizes for System with 128 GiB RAM

RAM Resource Owner	Total Size (GiB)
Tenant VM(1)	118
OVS	2
vCIC	-
vFuel	-
Host OS(2)	8(2)
Total (with host OS)	128

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.
(2) If ScaleIO is installed, SDC uses 0.05 GiB of the RAM configured for the host OS.

Table 9 Total Memory Sizes for System with 64 GiB RAM

RAM Resource Owner	Total Size (GiB)
Tenant VM(1)	54
OVS	2
vCIC	-
vFuel	-

RAM Resource Owner	Total Size (GiB)
Host OS[2]	8[2]
Total (with host OS)	64

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.
(2) If ScaleIO is installed, SDC uses 0.05 GiB of the RAM configured for the host OS.

3.5.3 Compute Server with vCIC

Table 10 and Table 11 specify the volumes allocated to the resource owners. Host OS is not included, so the tables contain 114 and 50 GiB, respectively.

Table 12 and Table 13 contain the total memory sizes.

Table 10 Memory Allocation for System with 128 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
Tenant VM[1]	1024	82	82
OVS	2	1024	2
vCIC	1024	30	30
vFuel	-	-	-

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

Table 11 Memory Allocation for System with 64 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
Tenant VM[1]	1024	18	18
OVS	2	1024	2
vCIC	1024	30	30
vFuel	-	-	-

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

Table 12 Total Memory Sizes for System with 128 GiB RAM

RAM Resource Owner	Total Size (GiB)
Tenant VM[1]	82

RAM Resource Owner	Total Size (GiB)
OVS	2
vCIC	30
vFuel	-
Host OS(2)	14(2)
Total (with host OS)	128

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.
(2) If ScaleIO is installed, SDC uses 0.05 GiB of the RAM configured for the host OS.

Table 13 Total Memory Sizes for System with 64 GiB RAM

RAM Resource Owner	Total Size (GiB)
Tenant VM(1)	18
OVS	2
vCIC	30
vFuel	-
Host OS(2)	14(2)
Total (with host OS)	64

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.
(2) If ScaleIO is installed, SDC uses 0.05 GiB of the RAM configured for the host OS.

3.5.4 Compute Server with vFuel

Table 14 and Table 15 specify the volumes allocated to the resource owners. Host OS is not included, so the tables contain 120 and 56 GiB, respectively.

Table 16 and Table 17 contain the total memory sizes.

Table 14 Memory Allocation for System with 128 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
Tenant VM(1)	1024	115	115
OVS	2	1024	2
vCIC	-	-	-
vFuel	1024	3	3

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

Table 15 Memory Allocation for System with 64 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
Tenant VM(1)	1024	51	51
OVS	2	1024	2
vCIC	-	-	-
vFuel	1024	3	3

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

Table 16 Total Memory Sizes for System with 128 GiB RAM

RAM Resource Owner	Total Size (GiB)
Tenant VM(1)	115
OVS	2
vCIC	-
vFuel	3
Host OS(2)	8(2)
Total (with host OS)	128

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.
(2) If ScaleIO is installed, SDC uses 0.05 GiB of the RAM configured for the host OS.

Table 17 Total Memory Sizes for System with 64 GiB RAM

RAM Resource Owner	Total Size (GiB)
Tenant VM(1)	51
OVS	2
vCIC	-
vFuel	3
Host OS(2)	8(2)
Total (with host OS)	64

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.
(2) If ScaleIO is installed, SDC uses 0.05 GiB of the RAM configured for the host OS.

3.5.5 Compute Server with vCIC and vFuel

Table 18 and Table 19 specify the volumes allocated to the resource owners. Host OS is not included, so the tables contain 114 and 50 GiB, respectively

Table 20 and Table 21 contain the total memory sizes.

Table 18 Memory Allocation for System with 128 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
Tenant VM(1)	1024	79	79
OVS	2	1024	2
vCIC	1024	30	30
vFuel	1024	3	3

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

Table 19 Memory Allocation for System with 64 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
Tenant VM(1)	1024	15	15
OVS	2	1024	2
vCIC	1024	30	30
vFuel	1024	3	3

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

Table 20 Total Memory Sizes for System with 128 GiB RAM

RAM Resource Owner	Total Size (GiB)
Tenant VM(1)	79
OVS	2
vCIC	30
vFuel	3
Host OS(2)	14(2)
Total (with host OS)	128

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.
(2) If ScaleIO is installed, SDC uses 0.05 GiB of the RAM configured for the host OS.

Table 21 Total Memory Sizes for System with 64 GiB RAM

RAM Resource Owner	Total Size (GiB)
Tenant VM(1)	15
OVS	2
vCIC	30
vFuel	3
Host OS(2)	14(2)
Total (with host OS)	64

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.
(2) If ScaleIO is installed, SDC uses 0.05 GiB of the RAM configured for the host OS.

3.5.6 ScaleIO Server with MDM/TB and SDS

Table 22 specifies the volumes allocated to the resource owners.

Table 23 contains the total memory sizes.

Table 22 Memory Allocation for System with 128 or 64 GiB RAM

RAM Resource Owner	Hugepage size (MiB)	Number of Hugepages (Count)	Total Size (GiB)
MDM/TB			0.5
SDS			0.5

Table 23 Total Memory Sizes for System with 128 or 64 GiB RAM

RAM Resource Owner	Total Size (GiB)
MDM/TB	0.5
SDS	0.5
Host OS	8
Total (with host OS)	9

3.6 Storage Configuration

This section describes storage implementations and disk requirements.

On BSP, the supported storage solutions are local storage and distributed storage.

3.6.1 Local Storage Disk Space

This section lists requirements on disk space.

- For Compute hosts based on GEP5-64, see Table 24.
- For Compute hosts based on GEP7-128, see Table 25.
- For vCIC hosts and Compute hosts based on GEP5-64-400 ,GEP5-64-1200, GEP7-128-X16 and GEP7-128-X, see Table 26.
- For vCIC, see Table 27.

The following are supported as Compute hosts:

- ROJ 208 866/5 GEP5-64
- ROJ 208 868/5 GEP5-64-400
- ROJ 208 867/5 GEP5-64-1200
- ROJ 208 841/7 GEP7-128-X
- ROJ 208 844/7 GEP7-128-X16

GEP5-64 has a 8 GiB disk, and GEP7-128 has a 16 GiB disk. CEE infrastructure use 7 GiB of those, and only 1 GiB is available for ephemeral storage for applications in the case of GEP5-64, and 9 GiB in the case of GEP7-128.. This small size is in practice only feasible for applications that run a RAM based file system and use anther compute host with disk for permanent storage as, for instance, done by IMS.

The disk space on GEP5-64 and GEP7-128 is not enough to store debug logs. Instead, the logs from Compute nodes are forwarded to the vCIC and stored in the vCIC. A different configuration can be used in BSP systems where all Compute hosts have a 400 GiB, 1200 GiB, or 1600 GiB disk, but such configurations are outside the scope of this document.

Only ROJ 208 867/5 GEP5-64-1200 and ROJ 208 844/7 GEP7-128-X16 are supported as hosts for vCIC.

Table 24 Compute Host with 8 GiB Disk, GEP5-64

Use	Compute Host
Root partition (host OS)	7 GiB
Logs and crash dumps	0 GiB
Remaining storage is for ephemeral storage for VMs	1 GiB

Table 25 Compute Host with 16 GiB Disk, GEP7-128

Use	**Compute Host**
Root partition (host OS)	7 GiB
Logs and crash dumps	0 GiB
Remaining storage is for ephemeral storage for VMs	9GiB

Table 26 vCIC Host and Compute Host, GEP boards

Use	**vCIC Host** **GEP5-64-1200** **GEP7-128-X16**	**Compute Host** **GEP5-64-400 (372 GiB) or GEP5-64-1200 (1116 GiB)** **GEP7-128-X**	**Note**
Root partition (host OS)	50 GiB	50 GiB	
Logs and crash dumps	40 GiB	40 GiB	
vCIC Total	554 GiB	-	vCIC backup is not included. See Table 27.
vCIC backup	52 GiB	-	
Host total without vFuel and Atlas	**696 GiB**	**90 GiB**	
vFuel	50 GiB	50	vFuel can be run on any Compute host and it uses disk from the ephemeral storage. **Limitation:** Only two of the following can be run on the same Compute host: • vCIC • vFuel • Atlas

Use	vCIC Host GEP5-64-1200 GEP7-128-X16	Compute Host GEP5-64-400 (372 GiB) or GEP5-64-1200 (1116 GiB) GEP7-128-X	Note
Atlas	130 GiB	130	Atlas can be run on any Compute host and it uses disk from the ephemeral storage. The disk for Atlas is allocated by the CEE. **Limitation:** Only two of the following can be run on the same Compute host: • vCIC • vFuel • Atlas
Remaining storage is used as ephemeral storage for VMs.	Dimensioned depending on application need.	Dimensioned depending on application need.	Valid for boot from image. Calculated from total disk reduced by used space.

Table 27 vCIC Disk Allocation

Note: This configuration is optimized for a BSP system equipped with 72 blades. Other configurations are possible, but outside the scope of this document.	
Use	**vCIC**
Root partition (host OS)	50 GiB
Logs and crash dumps	224 GiB
Database for OpenStack and Zabbix (MySQL)	90 GiB
Database for Ceilometer (MongoDB)	70 GiB

Glance repository in Swift	100 GiB
vCIC sum	554 GiB

3.6.2 Disk Requirements for Atlas

When Virtual Machine (VM) images are loaded to Atlas as part of an `.ova` file, the image is temporarily stored in ephemeral storage in Atlas. To support loading of large images, the recommendation is to use 120 GiB for the Atlas ephemeral storage.

The Atlas VM occupies 120 GiB of the local disk on the compute node where it is running.

To reduce the disk allocated to Atlas, the size of the ephemeral disk can be reduced from 120 GiB to a minimum of 10 GiB. Since 30% of the ephemeral disk in Altas is used as temporary storage for .ova files, the size of the ephemeral disk needs to be adjusted according to the size of .ova files to be loaded. Using a reduced disk size of 10 GiB implies that it can be impossible to load .ova files that contain images larger than 3 GiB.

3.6.3 Disk Requirements for Nova Snapshots

Nova snapshots are stored in the `/var/lib/glance` partition of CIC nodes.

There are certain disk requirements for the Nova snapshots to work. Depending on the requirements and frequency on Nova snapshots, the system must be dimensioned with free disk space, according to the following guidelines:

- Disk partition `/var/lib/nova` in the compute host where the VM is hosted, must have **at least** double the space of the snapshot/VM size, for a successful Nova snapshot. The reason is that the snapshot will be first extracted locally in the compute node before it is uploaded to the Glance/Swift store.
- The disk space needed in the `/var/lib/nova` partition of the compute disk must have free space **at least** twice the size of VMs root disk. The reason is that during the extraction of the snapshot, first the delta of the VM disk will be extracted, after which the complete disk will be extracted.
- Disk partition `/var/lib/glance` in each CIC node must have free space **at least** equal to the root disk size of the VM, in order to accommodate the snapshot.

3.6.4 Disk Requirements for Distributed Storage (ScaleIO)

Distributed Storage is optional. It can be used as the back end for Cinder.

The following requirements must be fulfilled:

- Needed disk space for ScaleIO component on server: 1 GB
- Minimum disk space to be added as device to one SDS: 100 GB (this must be a physical disk)
- Minimum number of SDSs: 3

4 Characteristics

This section describes the system characteristics of CEE.

4.1 General System Limits

For the list of system limits, see Table 28.

Table 28 General System Limits

Slogan	Limit
RAM used by the infrastructure(1) See Section 3.5 on page 10 for more information.	Compute Host: • Host OS: 8 GiB • OVS: 2 GiB • vFuel (if present): 3 GiB
	vCIC Host: • Host OS: 14 GiB • OVS: 2 GiB • vCIC: 30 GiB • vFuel (if present): 3 GiB

Slogan	Limit
Number of hosts (blades)	In the certified configuration 3 GEP5-64-1200 is used by the CEE infrastructure. Additional blades are needed for the VNF. CEE supports 3–72 blades.
Number of cores occupied by infrastructure	In the certified configuration, the GEPs used as vCIC, vFuel and Atlas hosts are fully allocated to the CEE infrastructure. On a Compute host with only tenant VMs: 1–2 cores are used by the host OS. The actual allocation depends on the application, and needs to be measured and dimensioned with the VNF in use. 1–4 cores are used for OVS. The actual allocation depend on the bandwidth needs from the VNF. See Section 4.4.1 on page 24 for dimensioning rules. All other cores are available for the VNF.

(1) If Atlas is used, it runs at one of the hosts, and it uses 4 GiB of the RAM configured for the tenant VMs.

4.2 Orchestration Interface

The system limits for orchestration are listed in Table 29.

Table 29 Orchestration Limits

Slogan	Limits
Number of tenants	The maximum number of supported tenants is 50.

4.3 Tenant Execution Environment

This section describes the tenant-related limits on the environment.

4.3.1 Performance

Performance limits are listed in Table 30.

Table 30 Tenant Execution Performance

Slogan	Limits
Oversubscription	CPU, memory, and disk overcommit are not supported.

4.3.2 Resiliency

Resiliency-related tenant limits are listed in Table 31.

Table 31 Tenant Execution Resiliency

Slogan	Limits
Execution environment resiliency	The execution environment resiliency is relying on VM evacuation. States not conserved in storage are lost. Each hypervisor instance is not redundant, and, apart from attached storage, assumed to be a knock-out unit.

4.4 Network

This section lists the limits on the network.

Neutron with VLAN segmentation is used.

4.4.1 Performance

The virtual switch (vSwitch) performance is measured by the packet rate (packets per second). The packet size has a very limited impact on the packet rate.

Note: Therefore, the forwarded amount of data (bit per second) increases if the packet size is increased.

Table 32 shows the throughput for PHY to VM traffic cases, and Table 33 for VM to VM cases. Table 34 provides dimensioning guidelines for specifying the amount of capacity that is safe to use. Table 35 describes vSwitch capacity for bandwidth-based scheduling.

Table 32 Measured Per Host Forwarding Capacity, 64 Byte Frames, PHY to VM

Slogan	Limits
Bidirectional traffic from PHY	One PMD core is allocated to CSS. One HT is used by CSS, the other is idle. Value = 3.70 Mpps
	One PMD core is allocated to CSS. CSS uses both HTs. Value = 4.64 Mpps
	Two PMD cores are allocated to CSS. One HT in each core is used by CSS, the other is idle. Value = 7.46 Mpps
	Two PMD cores are allocated to CSS. CSS uses both HTs in each core. Value = 5.02 Mpps
	Four cores allocated to CSS. One HT in each core is used by CSS, the other is idle. Value = 7.27 Mpps

Table 33 Measured Guest VM Delivery Forwarding Capacity, 64 Byte Frames, VM to VM Intrahost Traffic

Slogan	Limits
Bidirectional traffic from VM to VM	One PMD core is allocated to CSS. One HT is used by CSS, the other is idle. Value = 3.10 Mpps
	One PMD core is allocated to CSS. CSS uses both HTs. Value = 4.63 Mpps
	Two PMD cores are allocated to CSS. One HT in each core is used by CSS, the other is idle. Value = 5.30 Mpps
	Two PMD cores are allocated to CSS. CSS uses both HTs in each core. One vNIC per VM. Value = 4.39 Mpps
	Four cores allocated to CSS. One HT in each core is used by CSS, the other is idle. One vNIC per VM. Value = 5.79 Mpps

Table 34 Dimensioning Capacity (Bidirectional Traffic)

Slogan	Limits
Total vSwitch capacity (bidirectional traffic)	The total vSwitch capacity to be used for dimensioning is 80% of the per host forwarding value above. It is different from the value of the ‘‘Measured per Host Forwarding Capacity’’, in order to take into account external effects impacting the deterministic behavior of the v Switch. The user can use a different value tuned for a specific system configuration, preferably based on measurements.
Per interface vSwitch capacity (bidirectional traffic)	The maximum dimensioning limit per interface is 80% of the per host forwarding value for one PMD core allocated to OVS, when the HT functionality is not used. If HT is used on the cores hosting OVS PMD threads, the value is 50% of the per host forwarding value. If the number of interfaces is not bigger than the number of PMD threads, the 2 core values can be used as base, but this is not a likely scenario. If more interfaces are configured than OVS-assigned PMD threads, the maximum dimensioning limit per interface is reduced by an additional factor: the number of interfaces divided by the number of OVS PMD threads, rounding any fraction to the next higher integer. Two examples: 7 interfaces and 3 OVS PMD threads => 7/3 = 2.33, round up => dividend is 3, which means that the capacity figure should be divided by 3; 9 interfaces and 3 OVS PMD threads => 9/3 = 3, there is no fraction so no round up => dividend is 3, which means that the capacity figure should be divided by 3. It is not recommended to change the per interface limit, even if measurements indicate a higher capacity, as the behavior is highly dependent on the automatic distribution of the interfaces over the PMD threads.

Table 35 Virtual Switch Capacity for Bandwidth-Based Scheduling

Slogan	Limits
vSwitch capacity for bandwidth-based scheduling	The value is used as maximum threshold for the vSwitch bidirectional throughput per host. It is used to configure bandwidth-based scheduling per host when installing CEE. It must be below or, at most, equal to ‘‘Total vSwitch capacity (bidirectional traffic)’’ detailed in Table 34.

4.4.2 Resiliency

Network resiliency is listed in Table 36.

Table 36 Network Resiliency

Slogan	Limits
Self-healing network	The network solution is self-healing, including network fault detection and automated failover.

4.4.3 Tenant Network Limitations

Limitations of the tenant network are listed in Table 37.

Table 37 Tenant Network Limitations

Slogan	Limits
Number of virtual networks	The theoretical aggregated maximum number of virtual tenant networks per CEE region is 4050. Since each Neutron network created consumes RAM in the vCIC, this theoretical maximum cannot be reached. The default configuration of RAM for vCIC allows 1000 networks. Additional memory is needed if more Neutron Networks are created. For rough estimations, consider that 100 Neutron networks with 1 subnet and 1 port for each cost about 2 GiB memory.
Number of vNICs per guest VM	The maximum number of vNICs per guest VM is 10 (+ 1 Trunk vNIC).
Number of Trunk vNIC attached vLANs	The number of Trunk vNIC attached vLANs is limited to 100.

Slogan	Limits
Number of vNICs per blade	CSS supports up to 128 vNICs per Compute host.
L2 Packet MTU	The L2 Packet MTU size is 2140 bytes for the BSP internal network. Jumbo frames are not supported on the external network due to limitations in CMX.

4.5 Storage

This section describes CEE characteristics on storage.

4.5.1 Limitations

On BSP, the supported storage solutions are local storage and distributed storage.

For tenants, ephemeral storage (non-persistent block storage) is supported on local disks of the compute hosts. Persistent block storage is not supported on BSP systems.

There is no support for any shared file system in CEE. For distributed storage, see separate sections.

Swift uses the local storage, and Swift is also supported on distributed storage ScaleIO.

Object storage through Swift is only used for the CEE infrastructure.

Management of VM images is supported by the OpenStack image service.

Only boot from image is supported for BSP.

Table 38 shows where data is stored.

Table 38 Storage Locations

Data	Storage Location
CEE infrastructure backups (incl. Fuel backups)	On local disks of vCIC hosts(1)
Ephemeral storage	On local disks of Compute hosts(1)

Data	Storage Location
vCIC storage (OpenStack infrastructure)	On local disks of vCIC hosts(1)
Core/crash dumps, logs	On local disks of all hosts

(1) Local storage is limited by the local, non-scalable disk capacity.

Data is stored on a local disk, and it is erased in case of disk failure or rollback from a failed update, meaning that the VM disappears. The application must be designed accordingly.

4.5.2 VM Migration with NoMigration Policy Set

When a VM has `No Migration` policy set and is booted from local storage, it will not be started again after a rollback since the `/var/lib/nova` partition is not preserved.

The ephemeral disk for the VM is stored on local disk of the compute node. If the compute node must be replaced (because of, for instance, HW failure), any change in the VM is lost.

Boot from Volume is not available in BSP configurations.

4.5.3 Resiliency

Storage resiliency characteristics are listed in Table 39.

Table 39 Storage Resiliency

Slogan	Characteristics
Swift storage resiliency	Swift storage is replicated over the local disks that run vCIC.(1)

(1) Glance uses Swift.

4.6 In-Service Performance

This section lists the characteristics on in-service performance.

Table 40 In-Service Performance

Slogan	Characteristics
Guest execution retainability	Guest execution is not interrupted at a virtual Infrastructure management cluster restart or update.

Slogan	Characteristics
Update availability	When the update is running, OpenStack API is unavailable for about a minute for each CIC node. During rollback, negative response is occasionally returned.
Restart availability	It is not possible to connect to the API during the restart. The applications are designed to handle this and will not time out during restart.

5 System Limitations

This section describes the system limitations in R6.

5.1 OpenStack Deviations

The major deviations from the OpenStack SW are:

- Floating IP
- Object Storage
- Live Migration
- Security Groups

See relevant API descriptions for more information about limitations.

Limitations, Listed in API Documents

See the following API documents for more information on limitations:

- *Atlas OVFT API*
- *In Service Performance Northbound API*
- *Fault Management Northbound API*
- *Performance Management Northbound API*
- *Preconfigured Key Performance Indicators*
- *OpenStack API Complete Reference*

5.2 SW Configurations and Options

This section describes SW configurations and options.

5.2.1 Allocation of Memory

CEE supports flavors that allocate VM memory aligned to n × 1 GiB memory when hugepages are enabled. There is a Nova patch that makes Nova aware of this. The hugepages are in chunks of 1 GiB memory, all of which is reserved to the VM even if less memory is asked for.

Each Neutron network created consumes RAM in the vCIC, and it influences the maximum number of virtual tenant networks. See Section 4.4.3 on page 27 for more information.

5.2.2 Allocation of vCPU

The vCPU is limited to even number of vCPUs, see Section 4.1 on page 22.

5.2.3 Collocation of vCIC, vFuel, and Atlas

Only two of the following can be run on the same Compute host: vCIC, vFuel, Atlas.

5.2.4 Number of Parallel Root Volume Operations

Nova in CEE supports about 500 parallel stop/detach root volume operations.

5.3 Not Supported

This section describes functionalities that are not supported. These are included here because they are not related to any specific configurations.

5.3.1 Dashboard Does Not Support Internet Explorer

The Dashboard does not support Internet Explorer because of the Ericsson Graphical User Interface (GUI) Software Development Kit (SDK).

5.4 Limitations and Workarounds

CM-HA operates in active-passive mode. If a Compute host containing a vCIC that runs the active CM-HA restars, the VM evacuation will not start within one minute. The evacuation only starts when the CM-HA process is moved

to another vCIC by Corosync, and the Compute unavailability is detected by the CM-HA.

5.5 Update Limitations

OpenStack API is not always available during update and rollback. When the update is running, the OpenStack API is unavailable for about a minute for each CIC node. During rollback, a negative response is occasionally returned. For more information, see Section 4.6 on page 29.

5.6 Hardware Platform Limitations

This section describes the limitations for Ericsson Blade Server Platform (BSP).

CEE R6 has the following limitations:

- Only the following GEP boards are supported, as specified in Table 41:
  - GEP5-64-1200
  - GEP5-64-400
  - GEP5-64
  - GEP7-128--X16
  - GEP7-128-X
- Boards with vCIC node must have disks.

  Note: Only GEP5-64-1200 and GEP7-128-X16 are supported as vCIC host in the certified configuration.
- Compute blades can be “diskless”, for example, GEP5-64 or GEP7-128-X. In this case, the 8 or 16 GiB flash disk is used as the local disk.
- If diskless compute blades are used, the available memory for ephemeral disk is small (about 1 GiB on GEP5-64 and 9 GiB on GEP7-128-X).
- Supported versions: BSP R9 or higher

  Note: For known BSP problems affecting CEE, consult the *Limitations and Workarounds for Cloud Execution Environment (CEE)*.

Table 41 Supported Hardware

Item	Status
Servers	
BSP[1]	EBS with the following: • ROJ 208 866/5, GEP5-64G • ROJ 208 868/5, GEP5-64-400 • ROJ 208 867/5, GEP5-64-1200 • ROJ 208 841/7, GEP7-128-X • ROJ 208 844/7, GEP7-128-X16 • SCXB3 – ROJ 208 395/1[2] • CMXB3 – ROJ 208 392/1[2]

(1) BSP means fimware + BIOS.
(2) CEE supports the hardware versions of the board that are supported by BSP. For more information, refer to BSP Hardware Baseline, Reference [3].

Reference List

[1] *BSP Technical Product Description*, 221 02-FGC 101 2255

[2] *Juniper Networks homepage, www.juniper.net*

[3] *BSP Hardware Baseline*, 1090-CRA 119 1772, available at Ericsson Support