Dell Technologies, VMware, and Mavenir 5G O-RAN Reference Architecture

Technical preview

July 2021

H18828

Abstract
This guide describes a high-level reference architecture that is being planned by Dell Technologies and its partners, Mavenir and VMware, for complete testing and release in the second half of 2021. This guide can be used as the basis for future commercial deployments of 5G Open RAN networks. It provides a reusable model to help CSPs accelerate the adoption of this new technology.
Contents

Introduction ................................................................................................................................... 4

RAN Virtualization (vRAN) overview ............................................................................................ 5

vRAN reference architecture ........................................................................................................ 5

Mavenir software stack components ........................................................................................... 9

Deployment models .................................................................................................................... 13

VMware Telco Cloud Platform overview .................................................................................... 15

VMware Telco Cloud Platform for RAN overview ...................................................................... 15

VMware Telco Cloud Automation overview ............................................................................... 16

VMware Telco Cloud Automation Overview .............................................................................. 17

VMware Telco Cloud Platform for RAN deployment architecture ............................................ 18

PTP ............................................................................................................................................... 20

RAN virtualization design ........................................................................................................... 21

Workload domains and vSphere cluster design .......................................................................... 22

Network topology and hardware configurations for vCU/vDU .................................................. 22

vCU and vDU server specifications ........................................................................................... 23

Conclusion................................................................................................................................... 30
Introduction

The telecom industry is facing challenging times to efficiently accommodate the exponential growth in network traffic. New use cases and scarce spectrum resources have placed tremendous pressure on Communication Service Providers (CSPs) to make the most efficient use of their allocated radio spectrum. To address increased demands, CSPs are seeking new ways to design flexible networks, dynamically scale network capacity, expand service coverage, rapidly deploy new services, and improve the overall user experience while reducing the Total Cost of Ownership (TCO) and Time to Market (TTM).

There is a need for a next-generation radio evolution to accelerate bringing new services to market and improve spectral efficiency to increase network capacity and decrease latency. This evolution is necessary to realize the next generation of major use cases such as Enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency (URLLC), and Massive Machine Type Communications (mMTC). Next generation radio starts with the 5G New Radio (NR) air interface—the new standard for 5G wireless technology that can provide higher performance, lower TCO, and greater flexibility and scalability.

In addition, previous 4G Centralized Radio Access Network (C-RAN) architectures have paved the way for a newer, more disaggregated RAN architecture built entirely on cloud-native principles. Network virtualization capabilities that had been applied in the Core Network domain can now be applied to the RAN domain, allowing for a new cloud-native Open RAN solution that features new capabilities and services.

Market developments are proving that the telco domain is ripe for change and innovation. Geopolitical concerns are restricting the number of vendors who are able to deploy 5G networks in certain regions, while several CSPs now state that the established vendor lock-in is restricting innovation and does not allow for more favorable horizontal market conditions. Reliance on a few select vendors creates market, financial, and systemic risks for the overall CSP landscape in many countries.

Open RAN addresses new CSP challenges by:

- Promoting open protocols and interfaces within the mobile network
- Enabling a transition away from a limited number of vendors promoting proprietary end-to-end solutions
- Opening the market to industry-leading solutions supported by numerous vendors
- Fully supporting either Virtualized RAN (VRAN) or Physical RAN (PHY RAN) solutions

To this end, the Open RAN (O-RAN) Alliance is creating the next level of RAN specifications beyond the 3rd Generation Partnership Project (3GPP), which standardized the LTE and 5G NR air interfaces.

O-RAN-based mobile networks improve the efficiency of RAN deployments and mobile network operations. O-RAN introduces cloud-scale economies and competition to the RAN market by enabling an open, multivendor RAN ecosystem. Those marketplace changes, combined with a more flexible RAN architecture made possible by virtualization, enable a faster TTM than was previously possible.
The new O-RAN standards enable a more competitive and vibrant RAN supplier ecosystem with faster innovation. These advancements offer CSPs the service differentiation they need to compete on a basis of services, instead of the current paradigm of competing solely based on coverage and consumer user speeds.

Dell Technologies, Mavenir, and VMware have partnered to produce this reference architecture guide to offer a technical preview of an upcoming validated reference architecture, planned for the second half of 2021. This solution will power a 5G Open RAN network that will enable the changes required by CSPs for the next generation of cutting-edge services.

RAN Virtualization (vRAN) overview

Network Function Virtualization (NFV) is a concept that has been developing in wireless networks for the past decade. Standardization in the European Telecommunications Standards Institute (ETSI) and the 3GPP allows for breaking away from tightly coupled proprietary hardware and software solutions and adopting virtualization software running on commercial off-the-shelf (COTS) hardware. These NFV solutions were first used in core networks and transport solutions but are now available in the RAN.

Cloud-based RAN deployment is one of the fundamental tenets of the O-RAN architecture. Operators are developing NFV requirements that enhance virtualization platforms in support of various RAN software protocol splits to enhance network performance and decrease TCO. For example, a high layer split between Packet Data Convergence Protocol (PDCP) and Radio Link Control (RLC), and a low layer split within PHY (Physical Layer) is now common in the cellular wireless industry.

The O-RAN Alliance leverages and verifies the performance of relevant open-source communities including Kubernetes (K8S) and OpenStack to design programmable hardware accelerators, real time processing, and virtualization technologies.

vRAN reference architecture

Open vRAN is a software-intensive, scalable web architecture with open interfaces that disaggregate the hardware and software. Disaggregation allows a multivendor solution that combines industry-leading products for each element of the network and enables the use of COTS hardware. This approach helps simplify deployment and operation of a CSP's network assets.

The following figure shows the reference architecture’s solution stack and high-level network design.
Mavenir provides the following components in the stack:

- Mavenir Centralized Management Service (mCMS)
- Smart Deployment as a Service (SDaaS)
- Analytics
- 5G and regular Centralized Unit-Control Plane (CU-CPs)
- 5G Distributed Unit (DU)
- Mavenir Telco Cloud Integration Layer (MTCIL)

VMware provides the Cloud Deployment Manager (CDM). The Cloud deployment manager is responsible for deploying CNFs.

The VMware and Mavenir software stack components are deployed on Open RAN optimized high-performance Dell EMC PowerEdge XR11/XR12 and R750 servers.

Descriptions of Mavenir components and interfaces are in the following sections.

The Mavenir 5G Open RAN solution is a fully containerized system that virtualizes the baseband unit functions into the Open Centralized Unit (O-CU) and the Open Distributed Unit (O-DU). Virtualizing the control plane and user plane functions allows the O-CU to be consolidated in one or more data centers on RAN-based open interfaces. These data centers can be distributed geographically, which results in significant CAPEX and OPEX savings.

Each cell-site can have multiple third-party O-RUs and antennas that are connected to Mavenir’s O-DU using an O-RAN 7.2x fronthaul interface over the enhanced Common Public Radio Interface (eCPRI). The O-DU can be located at the site, or in a local data center (LDC), serving multiple O-RUs. Centralizing the DU instead of distributing the DU to the cell site results in several performance and cost advantages.
The Mavenir 5G Open RAN solution offers the following advantages.

- **Scalable architecture**—Provides a software architecture in O-CU-CP, O-CU-UP, and gNB O-DU that scales with the available cores. The O-CU-CP scales in terms of user equipment (UE), the O-CU-UP scales in terms of bearers, and the O-DU scales in terms of carriers and UEs.

- **Network densification**—Provides a flexible radio infrastructure that adapts to usage patterns for improved coverage and enhanced user experiences. Wireless service providers can deliver extra capacity and coverage in rural or high data traffic areas, for an enhanced user experience.

- **Scalability**—Guarantees scaling, availability, and performance in a cloud-native, fully virtualized environment. Each network function can be dimensioned independently based on the services offered by that network function.

- **Adaptability**—Enables vRAN disaggregation (3GPP + O-RAN) and continuous adaptation for interoperability using open standard interfaces.

- **Simplicity**—Simplifies operational complexity leading to zero-touch whitebox or COTS deployments.

**Split architecture** Since the earliest phases of 5G Radio Access standardization, there has been a push to decouple the radio access network by breaking off into Radio Unit (RU), Distributed Unit (DU), and Centralized Unit (CU) functions. How this decoupling takes places and the decision about which unit controls certain operations is known in the industry as functional splits.

The Mavenir 5G Open RAN architecture supports both Higher Layer Split (HLS) (3GPP Option 2), and the O-RAN Alliance Low Layer Split (LLS) 7.2x. The following diagram shows the possible split options in Open RAN and supported splits:
Figure 3. Protocol split architectures

Split option 2

Split option 2 is between PDCP Packet Data Convergence Protocol (PDCP) and RLC Radio Link Control (RLC). Since only the PDCP and above layers are centralized, split option 2 is less susceptible to network latencies and can operate under less stringent midhaul conditions.

Benefits of higher layer split option 2 include:

- Less aggressive midhaul requirements, as the scheduler is closer to the cell site
- Standard interface (F1) for NR
- Support for longer latencies and jitter (including wireless midhaul)
- Joint optimization for Hybrid Automatic Repeat Request (HARQ) and Automatic Repeat Request (ARQ) for reliable communication over the radio link in challenging low-latency scenarios
- Compatible with dual connectivity architecture being developed for NR
- Allows a simple bearer to split and adapt RLC segmentation

Split option 7.2x

In this option, the PHY layer’s functional modules are distributed between Low-PHY and High-PHY based on Open RAN specifications. The Split 7.2x objectives are:

- Minimize impact on transport bandwidth while maximizing virtualization in gNB CU and gNB DU
- Enable simple, low-cost RRU designs for wide adoption
- Eliminate performance loss compared to integrated solutions with ideal fronthaul
- Eliminate limitation on receiver architecture for performance
- Eliminate redesign for NR as opposed to LTE
- Provide increased scalability with fixed-rate streaming interface – transport data rate scales with traffic and bandwidth
- Centralize scheduling
- Support advanced signal processing such as UL compression
The diagram below illustrates Mavenir's Open RAN architecture’s functional split among O-CU, O-DU, and O-RU with 7.2x Fronthaul Split support.

**Mavenir software stack components**

**mCMS**

Mavenir Centralized Management System (mCMS) is a standard element management system (EMS) that manages all configurations and faults of the Cloud-Native Network Functions (CNF). It also plays a critical role in initializing the CNFs.

The key architectural features of mCMS are:

- Web-based UI
- Configuration management
- Inventory management
- Performance and fault management
- Provides a network topology:
  - Describes CNF and link status
  - Enables northbound interfaces

The following diagram shows the architecture view of the mCMS managing CU and DU applications:
SDaaS

The Mavenir 5G RAN solution includes Smart Deployment as a Service (SDaaS). SDaaS provides the following services:

- Plug and play manager for all Physical Network Functions (PNF) and third-party nodes
- Fault, Configuration, Accounting, Performance, Security (FCAPS) for all PNF and third-party nodes
- Inventory management for all PNF and third-party nodes
- Software upgrade management for all PNF and third-party nodes
- Network Function (NF) selection

Analytics

The components that comprise the Mavenir Observability Framework (OBF) solution are shown in the diagram below.
The components of the Analytics solution provide the following features:

- **Performance management (PM)**
- **Fault management (FM)**
- **Log management**

**Performance Management (PM)**

The Analytics solution uses Metricbeat to load the multiple time-series of raw counter data into ElasticSearch, and then performs transformations on the raw data to produce Key Performance Indicators (KPIs) that are displayed to the operator on the Kibana dashboard.

Mavenir Analytics is consolidating data into reliable, scalable Kafka streaming platform. It uses Logstash and Beats to normalize data if needed. Avro data format and stateless streaming jobs are used. Since Kafka data bus is designed to keep multiple uses of the same data independent of each other, the default northbound integration of Analytics is leveraging this functionality.

**Fault Management (FM)**

The fault events are generated and sent through the platform layer in VES format and are ingested and stored to ElasticSearch. From ElasticSearch, they can be retrieved with queries for correlation against KPIs and logs.

**Log Management**

During normal operations, the ASCII and binary logs are written to the local storage. The application is then responsible for file rotation. FluentBit currently supports a plug-in that allows binary and ASCII transfer. FluentBit stores binary and crash logs on remote storage. A binary scraper and parser provides on-demand capabilities that can pull binary and crash files from remote storage or local storage, parse them, and store them in ElasticSearch.
Mavenir software stack components

**5G CU-CP CNF**  
The central unit control plane (CU-CP) hosts RRC and the control-plane part of the PDCP protocol. The CU-CP terminates the E1 interface connected with the central unit user plane (CU-UP) and the F1-C interface connected with the distributed unit (DU).

**5G CU-UP CNF**  
CU-UP hosts the user-plane part of the PDCP and the Service Data Adaptation Protocols (SDAP). The CU-UP terminates the E1 interface connected with the CU-CP and the F1-U interface connected with the DU.

**5G DU CNF**  
The DU handles the lower layers of the baseband processing up through the PDCP layer of the protocol stack.

**Mavenir Telco Cloud Integration Layer (MTCIL)**  
MTCIL is a layer in the Mavenir Webscale Platform (MWP) that provides 3GPP and cloud-native compliant northbound and southbound interfaces for OSS, ORAN SMO, and PaaS integrations. It delivers the high levels of reliability, agility, and scalability that are expected of distributed and cloud-native systems.

The CU and DU containerized applications are hosted on the VMware PaaS and CaaS layers with Mavenir’s MTCIL on top of it. MTCIL is part of Mavenir Web-scale Platform (MWP), as is the Mavenir Cloud Platform layer. For the 5G Open RAN solution, Mavenir leverages and uses MTCIL of MWP. The PaaS and CaaS layers are from VMware.

The MCTIL layer provides these high-level functions:

- Exposes 3GPP and Cloud Native-compliant North and Southbound interfaces for OSS, ORAN SMO, and PaaS integrations.
- Exposes APIs and CNF packaging procedures using open interfaces and tools.
- Allows third party network functions to use these interfaces to enable Telco NF capabilities.
- Delivers the high level of reliability, agility, and scalability expected of distributed and cloud-native systems.

The diagram below shows an architectural view of the platform:
Deployment models

The Mavenir virtualized Open RAN solution supports the following deployment scenarios:

- DU and CU at the site
- DU (at the site with CU) is deployed remotely, while CU-CP and CU-UP are deployed at the same data center/cloud location
- DU at the Site with Control and User Plane Separation (CUPS), which means that CU-CP and CU-UP are deployed in different data center/cloud locations
- DU, CU-CP, and CU-UP software are deployed at separate data center/cloud locations
- DU and CU software are deployed at the same data center/cloud location

Of the five above deployment scenarios, the last four are projected to be the main scenarios.

The following figure depicts all the deployment scenarios supported by the Mavenir 5G O-RAN solution.
### Deployment models

**Figure 8. 5G RAN deployment models**

The following figure depicts the data center deployment model (urban eMBB use case) in which DUs are deployed in LDC, and CU-CP and CU-UP are deployed in RDC.

**Figure 9. Datacenter deployment model**
VMware Telco Cloud Platform overview

VMware Telco Cloud Platform is a modernization solution that deploys cloud-native and virtual network function consistently, at web-scale speed, and without disruption. VMware Telco Cloud Platform is a cloud-native platform that empowers CSPs to manage VNFs and CNFs across the core, far edge (RAN), enterprise edge, and cloud with efficiency, scalability, and agility. Telco Cloud Platform provides the framework to deploy and manage VNFs and CNFs quickly and efficiently across distributed 5G networks. You can run VNFs and CNFs from dozens of vendors, on any cloud, with holistic visibility, orchestration, and operational consistency.

VMware Telco Cloud Platform for RAN overview

VMware Telco Cloud Platform for RAN is a cloud-native RAN solution that is designed specifically for running RAN functions. It provides the RAN modernization path, evolving from legacy Radio Access Network (RAN) to virtualized RAN (vRAN) to Open RAN. It transforms the RAN into a 5G multi-services hub “mini cloud”, enabling CSPs to monetize their RAN investments.

Telco Cloud Platform RAN is designed to meet the performance and latency requirements inherent to RAN workloads:

- Enables CSPs to run virtualized Baseband functions that include virtualized Distributed Units (vDUs) and virtualized Central Units (vCUs).
- Simplifies CSPs’ operations consistently across distributed vRAN sites with centralized cloud-first automation, while reducing the Operating Expense (OpEx).
VMware Telco Cloud Automation overview

- Provides operational consistency by removing business uncertainties and reduces ballooning costs associated with 5G deployment.
- Enables CSPs to accelerate innovation speed, deploy 5G services fast, and scale the services as customers’ demands increase.

Telco Cloud Platform RAN is powered by field-proven virtualization, carrier-grade Container-as-a-Service (CaaS), and multi-layer automation that are consistent with 5G core and edge offerings. This end-to-end consistency, which is achieved by the coherent underpinning platform across 5G networks, enables CSPs to provision 5G services customized for different enterprise and consumer markets while providing unparalleled operational efficiency.

VMware Telco Cloud Automation overview

VMware Telco Cloud Automation (TCA) is a unified orchestrator. It onboards and orchestrates workloads seamlessly from VM and container-based infrastructures. It distributes workloads from the core to the edge and from private to public clouds for unified orchestration.

TCA accelerates time-to-market for network functions and services while igniting operational agility through unified automation, across any network and any cloud. It applies an automated, cloud-first approach that streamlines the CSP’s orchestration journey with native integration to VMware Telco Cloud Infrastructure.

TCA enables multi-cloud placement, easing workload instantiation and mobility from the network core to edge and from private to public clouds. It also offers standards-driven modular components to integrate any multi-vendor MANO architecture. VMware further enhances interoperability by expanding partner network function certification using the VMware Ready for Telco Cloud program. With simplified and certified interoperability, CSPs can now leverage industry-leading solutions and reduce their risks.

VMware Telco Cloud Platform (TCP) is a common platform spanning Core and RAN functionality. This common platform self-tunes automatically depending on the workload deployed through VMware Telco Cloud Automation. To deploy all VNF/CNFs from 5G Core to RAN, the same automation platform, operational tools, and CaaS layer based on VMware Tanzu are used.
TCP RAN is a new compute workload domain that spans from the central or regional data center to cell sites. The management components of this compute workload domain reside in the management cluster inside the RDC management domain. Within this compute workload domain, VMware ESXi hosts are managed as single node hosts and distributed across thousands of cell sites. This distributed architecture also applies to Kubernetes. Kubernetes cluster management is centralized, and workload VMs are distributed to respective cell sites.

Each cell site compute workload domain can support up to 128 cell site locations because of the 128 virtual datacenter limit per vCenter. Telco Cloud Automation dedicates one virtual datacenter for each cell site host.

**Virtual platform infrastructure**

The virtual infrastructure is the foundation of the platform. It contains the software-defined infrastructure, software-defined networking, and software-defined storage.

**Compute workload domain**

With TCP RAN architecture, you can deploy and extend RAN applications into regional data center and cell sites. The compute workload domain that you deploy in the regional data center (RDC) through Telco Cloud Automation can contain multiple cell site vSphere hosts, each running the user workloads to enable the RAN solution. These cell site hosts can start with just one ESXi host, and can then scale depending on the resource and availability requirements of the solution being deployed.

**VMware Telco Cloud Automation Overview**

VMware Telco Cloud Automation (TCA) is a unified orchestrator. It onboards and orchestrates workloads seamlessly from VM and container-based infrastructures. It distributes workloads from the core to the edge and from private to public clouds for unified orchestration.
CSPs are transitioning from physical to cloud networks to gain operational agility, network resiliency, and low operating costs. This shift marks a radical departure from the traditional single-purpose hardware appliance model, especially as CSPs must now design and operate services across a web of data centers—bridging physical and virtual ecosystems—while enabling interoperability across competing vendors.

Due to the complexity of coordinating network functions and managing multiple services, CSPs require an automated approach that removes complexity and error-prone manual processes.

To address these challenges and improve operational efficiency, CSPs are turning to VMware TCA, as shown in the following figure.

**Figure 12. VMware TCA**

**VMware Telco Cloud Platform for RAN deployment architecture**

The architecture of the TCP RAN solution shows the management and compute workload placement between the regional data center site and the cell site. The following figure shows an end-to-end deployment model of TCP 5G Core and RAN to understand how they work together.

In the Regional Data Center (RDC) site, vSphere administrators deploy and configure one management workload domain and one or more cell site compute workload domains. The cell-site compute workload domain dedicates a vCenter server to manage both RDC and cell site vSphere hosts and workloads.
This deployment model includes two RDCs and one cell site as part of the VMware Telco Cloud Platform RAN solution architecture, and additional components as described below:

- The RDCs consist of one management workload domain and one cell site compute workload domain.
- The management workload domain hosts a dedicated vCenter and manages all the SDDC management and operational management components such as vCenter Server, NSX, vRealize Operations, vRealize Log Insight, and so on.
- VMware NSX is deployed as part of 5G Core components. It is used only for workloads that are specific to telco management at the RDC. NSX is not used for cell site compute workload domain functions.
- The compute workload domain hosts a dedicated compute vCenter server with one vSphere cluster such as WLD1-Cluster1 and multiple cell site hosts.
- In the compute vCenter server, WLD1-Cluster1 is hosted at the RDC and standalone hosts are deployed at cell site locations.
PTP

- A dedicated vSphere Distributed Switch (VDS) is used for both RDC vSphere cluster (WLD1-Cluster1) and the cell site host.
- Kubernetes cluster components such as control plane nodes are deployed at a vSphere cluster, such as WLD1-Cluster1, in the RDC.
- Kubernetes worker nodes are deployed at both the RDC and at cell site locations to support the telco CNF workloads such as vCU and vDU in a geographically distributed manner.

PTP

Precision Time Protocol (PTP) delivers time synchronization in various telco applications and environments. It is defined in the IEEE 1588-2008 standard. PTP helps issue accurate time and frequency over telecommunication mobile networks. Precise timekeeping is a key attribute for telco applications. It allows these applications to accurately construct the precise sequence of events that occurred or occur in real-time. PTP requires each ESXi node in the TCP RAN solution to be time-synchronized.

Note: The precision of a clock describes how consistent its time and frequency are, relative to a reference time source, when measured repeatedly. The distinction between precision and accuracy is subtle but important.

PTP profiles

A profile is a set of specific PTP configuration options that are selected to meet the requirements of telco RAN applications. PTP lets you define various profiles so that you can use PTP in different scenarios.

PTP traffic

If the network carrying PTP consists of a non-PTP-aware switch in the pathway between the Grandmaster and Follower clocks, the switch handles PTP as any other data traffic, affecting the PTP accuracy. In this case, use a proper Quality-of-Service (QoS) configuration for network delivery to prioritize PTP traffic over all other traffic.

PTP grandmaster clocks

Networks can be distributed geographically across different locations that include a central data center, regional data center, and cell sites that are connected over Wide Area Networks (WAN). Varying latencies across the WAN links can compromise PTP accuracy. In these circumstances, use different PTP Grandmaster clocks in each site and do not extend PTP across these sites.

The following guidelines apply for the PTP sources:

- The PTP sources such as Telco Grandmaster Clock (T-GM) must be reachable by all the components in the TCP RAN solution.
- Use the G.8275.1 PTP profile for accurate time synchronization for RAN applications. ITU–T G.8275.1 defines the PTP profile for network scenarios with the full-timing support, which means that all the intermediate switches support Boundary Clock functionality (BC).
- For the cell site locations, each host must have at least three dedicated physical ports: one port for PTP time synchronization and two ports for workloads.
The following figure shows the physical NIC placement for PTP on an ESXi host:

![PTP design for ESXi host at cell site location](image)

**RAN virtualization design**

The TCP RAN solution architecture consists of a centralized management domain along with compute workload domains to support the required RAN workloads in RDCs and cell sites.

Before you deploy a cell site host, deploy a management domain and a compute workload domain at your RDC using Telco Cloud Automation. The following section describes the VMware SDDC design within the construct of a management workload domain and a compute workload domain, and how these domains are related in this RAN design.

The management domain contains a single vSphere cluster called the management cluster. The management cluster hosts the VMs that manage the solution. This cluster is crucial for the management and monitoring of the solution. This highly available deployment ensures that the management and monitoring services are always available centrally at the RDC.

**Table 1. vSphere management components**

<table>
<thead>
<tr>
<th>Management domain components</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management vCenter Server</td>
<td>Manages the Management Domain.</td>
</tr>
<tr>
<td>Compute vCenter Server</td>
<td>Manages the Compute Workload Domain.</td>
</tr>
<tr>
<td>NSX Manager</td>
<td>Three instances of NSX Manager are used in a cluster.</td>
</tr>
<tr>
<td>Note: NSX is not used at the RAN site such as the cell site location.</td>
<td></td>
</tr>
<tr>
<td>vRealize Suite Standard</td>
<td>Includes vRealize Log Insight and vRealize Operations Manager.</td>
</tr>
<tr>
<td>vRealize Network Insight</td>
<td>Communicates with the vCenter Server and NSX Manager instances to collect metrics that are presented through various dashboards and views.</td>
</tr>
<tr>
<td>vRealize Orchestrator</td>
<td>Workflow engine, fully integrated with Telco Cloud Automation</td>
</tr>
<tr>
<td>Telco Cloud Automation</td>
<td>Includes TCA Manager and TCA-CPs.</td>
</tr>
</tbody>
</table>
### Management domain components

<table>
<thead>
<tr>
<th>Management domain components</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>VMware Tanzu Basic for RAN</td>
<td>Creates workload clusters in the compute workload domain.</td>
</tr>
</tbody>
</table>

**Note:** In the TCP RAN solution, the management domain is hosted at the RDC.

---

## Workload domains and vSphere cluster design

The compute workload domain can contain multiple vSphere clusters at the RDC. These clusters can contain a minimum of three ESXi hosts and a maximum of 64 hosts (when using vSAN), depending on the resource and availability requirements of the solution being deployed at the RDC.

The cell site host, which is designed to onboard CNFs on a single node ESXi host, is part of the compute workload domain. This host provides a Kubernetes workload cluster where CNFs such as vCU and vDU workloads are placed.

Each compute workload domain can support a maximum of 2500 ESXi hosts and 45,000 VMs in combination with RDC cluster and Cell Site hosts. If you use other management and monitoring tools, the vCenter maximums do not apply and the actual number of ESXi hosts and VMs per workload domain might be less.

---

## Workload domains and vSphere cluster design

The vCenter Server design for RAN includes the design for all the vCenter Server instances. For this design, determine the number of instances, their sizes, networking configuration, vSphere cluster layout, redundancy, and security configuration.

vCenter Server is deployed at Regional Data Center, and it manages all the Cell Site hosts. It is critical to design vCenter appropriately before onboarding the Cell Site hosts and RAN applications.

A vCenter Server deployment can consist of two or more vCenter Server instances according to the scale, number of VMs, and continuity requirements for your environment.

You must protect the vCenter Server system because it is the central point of management and monitoring. You can protect vCenter Server according to the maximum downtime tolerated. Use the following methods to protect the vCenter Server instances:

- Automated protection using vSphere HA
- Automated protection using vCenter Server HA

---

## Network topology and hardware configurations for vCU/vDU

The following diagram shows the high-level network topology for this technical preview version of the reference architecture. Dell EMC PowerEdge R750 CUs reside at the RDC and NDC locations, the Dell EMC PowerEdge XR12 and Dell EMC PowerEdge R750 CUs reside at the LDC locations, while the Dell EMC PowerEdge XR11 DUs reside at the cell sites (distributed DU model). Midhaul (F1 interface) connectivity is provided by an MPLS backbone transport network.
vCU and vDU server specifications

This section describes examples of servers on which this solution can be deployed. Although this reference architecture does not mandate server specifications for deploying Open RAN, the intent is to show a set of reusable examples which have been tested.

The Dell EMC PowerEdge R750, powered by 3rd Generation Intel Xeon Scalable processors, is a rack server that optimizes application performance and acceleration. The R750 is a dual-socket 2U rack server that delivers outstanding performance for the most demanding workloads. It supports 8 channels of memory per CPU, and up to 32 DDR4 DIMMs at 3200 MT/s speeds. In addition, to address substantial throughput improvements, the PowerEdge R750 supports PCIe Gen 4 and up to 24 NVMe drives with improved air-cooling features and optional Direct Liquid Cooling to support increasing power and thermal requirements. These features make the R750 an ideal server for data center standardization on a wide range of workloads including database and analytics, high performance computing (HPC), traditional corporate IT, virtual desktop infrastructure, and AI/ML environments that require performance, extensive storage, and GPU support.
## Table 2. Dell EMC PowerEdge R750 Server specifications

<table>
<thead>
<tr>
<th>Category</th>
<th>PowerEdge R750 features</th>
</tr>
</thead>
</table>
| CPU               | Up to two 3rd Generation Intel® Xeon® Scalable processors with up to 40 cores per processor  
Support for up to two 270 W processors  
Direct Liquid Cooling support |
| Accelerators      | Two 300 W (DW) or six 75 W (SW) |
| Memory            | Up to 32 DDR4 RDIMMs or LRDIMMs (8 TB + Optane Persistent Memory 200 Series)  
Optane Persistent Memory 200 Series (Barlow Pass): Yes  
NVDIMM: No  
DIMM Speed: Up to 3200 MT/s corrected |
| Storage           | **Front bays:**  
Up to 12 3.5-inch SAS/SATA (HDD/SSD) max 192 TB  
Up to 8 2.5-inch NVMe (SSD) max 122.88 TB  
Up to 16 2.5-inch SAS/SATA/NVMe (HDD/SSD) max 245.76 TB  
Up to 24 2.5-inch SAS/SATA/NVMe (HDD/SSD) max 368.84 TB  
**Rear bays:**  
Up to 2 2.5-inch SAS/SATA/NVMe (HDD/SSD) max 30.72 TB  
Up to 4 2.5-inch SAS/SATA/NVMe (HDD/SSD) max 61.44 TB |
| Storage Controller| Internal controllers: PERC H745, HBA355I, S150, H345, H755, or H755N  
Boot Optimized Storage Subsystem (BOSS-S2): HW RAID, either two 240 GB M.2 SSDs or one 480 GB  
External PERC (RAID): PERC H840, HBA355E |
| Network           | Two 1 GbE LOM  
1 OCP 3.0 (x8 PCIe lanes) |
| PCIe Slots        | Up to 8 PCIe Gen4 slots (up to 6 x16) with support for SNAP I/O modules |
| Integrated Ports  | **Front Ports:**  
1 Dedicated iDRAC Direct micro-USB  
1 USB 2.0  
1 VGA  
**Rear Ports:**  
1 USB 2.0  
1 Serial (optional)  
1 USB 3.0  
2 RJ-45  
1 VGA (optional for liquid cooling configuration) |
| Systems Management| iDRAC9  
iDRAC Service Module  
iDRAC Direct  
Quick Sync 2 wireless module |
<table>
<thead>
<tr>
<th>Category</th>
<th>PowerEdge R750 features</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Availability</td>
<td>Hot Plug redundant drives, three-tiered Hot Plug Fans, PSU, IDSDM, BOSS-S2 (2 x M.2)</td>
</tr>
<tr>
<td>Dimensions</td>
<td>Height – 86.8 mm (3.41 inches)</td>
</tr>
<tr>
<td></td>
<td>Width – 482 mm (18.97 inches)</td>
</tr>
<tr>
<td></td>
<td>Depth – 758.3 mm (29.85 inches) without bezel</td>
</tr>
<tr>
<td></td>
<td>772.14 mm (30.39 inches) with bezel</td>
</tr>
<tr>
<td>Form Factor</td>
<td>2U rack server</td>
</tr>
</tbody>
</table>

The Dell EMC PowerEdge XR11 server, which is powered by 3rd Generation Intel Xeon Scalable processors, is a high-performance, high-capacity server for demanding workloads at the edge. This is one of Dell Technologies’ latest 1U socket servers designed to run complex workloads using highly scalable memory, I/O, and network options, with up to eight DDR4 DIMMs, up to three PCIe Gen4 enabled expansion slots, up to four storage bays, and four integrated 25 GbE LAN on Motherboard (LoM) ports. It has a reduced form factor and is NEBS3 compliant, which makes it ideal for more challenging deployment models at the far edge, where space and environmental conditions become more demanding.

Figure 17. Dell EMC PowerEdge XR11 Server
**Table 3. Dell EMC PowerEdge XR11 Server specifications**

<table>
<thead>
<tr>
<th>Category</th>
<th>PowerEdge XR11 features</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>One 3rd Generation Intel Xeon Scalable processor with up to 36 cores</td>
</tr>
<tr>
<td></td>
<td>Support for one 225 W processor</td>
</tr>
<tr>
<td>Accelerators</td>
<td>Support PCIe Gen4 based Nvidia cards</td>
</tr>
<tr>
<td></td>
<td>Up to 2 T4</td>
</tr>
<tr>
<td>Memory</td>
<td>Optane Persistent Memory 200 Series</td>
</tr>
<tr>
<td></td>
<td>Up to 8 DDR4 RDIMMs or LRDIMMs (1DPC)</td>
</tr>
<tr>
<td>Storage</td>
<td>Up to four 2.5&quot; 12 Gb SAS, 6 Gb SATA, or NVMe SSDs</td>
</tr>
<tr>
<td></td>
<td>Four NVMe (direct or with RAID controller)</td>
</tr>
<tr>
<td></td>
<td>Optional BOSS 1.0 for boot</td>
</tr>
<tr>
<td>Storage Controller</td>
<td>PERC 10.5: H345</td>
</tr>
<tr>
<td></td>
<td>PERC 11: HBA355i, H755P</td>
</tr>
<tr>
<td></td>
<td>External: H840, HBA355e</td>
</tr>
<tr>
<td></td>
<td>Software RAID: S150</td>
</tr>
<tr>
<td>Network</td>
<td>Integrated LOM, four 25 GbE SFP+</td>
</tr>
<tr>
<td>PCIe slots</td>
<td>1 x8 PCIe Gen4 LP + 2 x16 PCIe Gen4 FH</td>
</tr>
<tr>
<td>Integrated Ports</td>
<td>Integrated LOM - (4x) 25 GbE SFP+</td>
</tr>
<tr>
<td></td>
<td>Serial Port</td>
</tr>
<tr>
<td></td>
<td>VGA</td>
</tr>
<tr>
<td></td>
<td><strong>USB XR11 with front access ports</strong></td>
</tr>
<tr>
<td></td>
<td>• Front: 1 standard USB 2.0, 1 micro; USB 2.0</td>
</tr>
<tr>
<td></td>
<td>• Rear: 1 standard USB 2.0, 1 standard USB 3.0; Internal: 1 standard USB 3.0</td>
</tr>
<tr>
<td></td>
<td><strong>USB XR11 with rear access ports</strong></td>
</tr>
<tr>
<td></td>
<td>• Front: 1 standard USB 2.0, 1 micro USB 2.0, 1 standard USB 3.0</td>
</tr>
<tr>
<td></td>
<td>• Rear: None; Internal: 1 standard USB 3.0</td>
</tr>
<tr>
<td>Systems Management</td>
<td>iDRAC9 Enterprise, Datacenter license options; OpenManage Enterprise and Plugins (Power Manager, SupportAssist, and Update Manager). iDRAC Direct</td>
</tr>
<tr>
<td>High Availability</td>
<td>Hot Plug Drives, Hot Plug Redundant Power Supplies, Redundant fans, Boot Optimized Storage Subsystem (BOSS 1.0)</td>
</tr>
<tr>
<td>Dimensions</td>
<td>H x W x D: XR11NAF – 42.8 mm x 434 mm x 400 mm; XR11RAF – 42.8 mm x 434 mm x 404 mm</td>
</tr>
<tr>
<td>Form Factor</td>
<td>1U Rack Server</td>
</tr>
</tbody>
</table>

The Dell EMC PowerEdge XR12 server, which is powered by 3rd Generation Intel Xeon Scalable processors, is a high-performance, high-capacity server for demanding workloads at the edge. This 2U socket server, one of Dell’s latest, is designed to run complex workloads using highly scalable memory, I/O, and network options, with up to eight DDR4 DIMMs, up to five PCIe Gen4 enabled expansion slots, up to six storage bays, and four integrated 25 GbE LAN on Motherboard (LoM) ports. It has a reduced form factor and is NEBS3 compliant, which makes it ideal for more challenging deployment.
models at the near or far edge, where space and environmental conditions become more demanding.

Figure 18. Dell EMC PowerEdge XR12 Server

Table 4. Dell EMC PowerEdge XR12 Server specifications

<table>
<thead>
<tr>
<th>Category</th>
<th>PowerEdge XR12 features</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>One 3rd Generation Intel Xeon Scalable processor with up to 36 cores</td>
</tr>
<tr>
<td></td>
<td>Support for up to 1 x 225 W processors</td>
</tr>
<tr>
<td>Accelerators</td>
<td>Support PCIe Gen4 based NVidia,</td>
</tr>
<tr>
<td></td>
<td>Up to 2 T4</td>
</tr>
<tr>
<td></td>
<td>Up to 2 A100</td>
</tr>
<tr>
<td></td>
<td>Up to 2 A40</td>
</tr>
<tr>
<td>Memory</td>
<td>Up to 8 DDR4 RDIMMs/LRDIMMs (8 TB + Optane Persistent Memory 200 Series)</td>
</tr>
<tr>
<td></td>
<td>Optane Persistent Memory 200: Yes</td>
</tr>
<tr>
<td>Storage</td>
<td>Up to six 2.5” 12 Gb SAS, 6 Gb SATA, NVMe SSDs</td>
</tr>
<tr>
<td></td>
<td>Up to six NVMe, either four NVMe directly attach or six with the H755P RAID controller</td>
</tr>
<tr>
<td></td>
<td>Optional BOSS 1.0 for boot</td>
</tr>
<tr>
<td>Storage Controller</td>
<td>PERC 10.5: H345</td>
</tr>
<tr>
<td></td>
<td>PERC 11: HBA355i, H755P</td>
</tr>
<tr>
<td></td>
<td>External: H840, HBA355e</td>
</tr>
<tr>
<td></td>
<td>Software RAID: S150</td>
</tr>
<tr>
<td>Network</td>
<td>Integrated LOM, four 25 GbE SFP+</td>
</tr>
<tr>
<td>PCIe slots</td>
<td>1 x8 PCIe Gen4 LP + 2 x16 PCIe Gen4 FH</td>
</tr>
<tr>
<td></td>
<td>or</td>
</tr>
<tr>
<td></td>
<td>1 x8 PCIe Gen4 LP + 4 x8 PCIe Gen4 FH</td>
</tr>
</tbody>
</table>
# vCU and vDU server specifications

<table>
<thead>
<tr>
<th>Category</th>
<th>PowerEdge XR12 features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated Ports</td>
<td>Integrated LOM, four 25 GbE SFP+</td>
</tr>
<tr>
<td></td>
<td>Serial Port</td>
</tr>
<tr>
<td></td>
<td>VGA</td>
</tr>
<tr>
<td><strong>USB XR12 with front access ports</strong> –</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Front: 1 standard USB 2.0, 1 micro USB 2.0</td>
</tr>
<tr>
<td></td>
<td>Rear: 1 standard USB 2.0, 1 standard USB 3.0</td>
</tr>
<tr>
<td></td>
<td>Internal: 1 standard USB 3.0</td>
</tr>
<tr>
<td><strong>USB XR12 with rear access ports</strong> –</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Front: 2 standard USB 2.0, 1 micro USB 2.0, 1 standard USB 3.0</td>
</tr>
<tr>
<td></td>
<td>Rear: None; Internal: 1 standard USB 3.0</td>
</tr>
<tr>
<td></td>
<td>Dedicated iDRAC network port</td>
</tr>
<tr>
<td>Systems Management</td>
<td>iDRAC9 Enterprise, Datacenter license options; OpenManage Enterprise and Plugins (Power Manager, SupportAssist, and Update Manager). iDRAC Direct</td>
</tr>
<tr>
<td>High Availability</td>
<td>Hot Plug Drives, Hot Plug Redundant Power Supplies, Redundant fans, Boot Optimized Storage Subsystem (BOSS 1.0)</td>
</tr>
<tr>
<td>Dimensions</td>
<td>H x W x D: XR12NAF – 86.8 mm x 434 mm x 400 mm; XR12RAF – 86.8 mm x 434 mm x 400 mm</td>
</tr>
<tr>
<td>Form Factor</td>
<td>2U Rack Server</td>
</tr>
</tbody>
</table>

The demanding physical layer processing required by 5G necessitates the use of an outboard PCIe hardware accelerator card. For the upcoming official release of the reference architecture described in this technical preview, Dell Technologies anticipates using Silicom’s Lisbon ACC100 Forward Error Correction (FEC) card. Silicom’s eASIC ACC100 FEC Accelerator server adapter is based on the Intel vRAN Dedicated Accelerator ACC100, an Intel eASIC Nextreme-3S device. The ACC100 is packaged in a 35 mm x 35 mm FC1156 package with 1.0 mm ball pitch Some balls are depopulated to facilitate break-out. The ACC100 provides 4G (Turbo) and 5G (LDPC) encode and decode over the PCIe interface.

![Hardware accelerator card](image)

Figure 19. Silicom Lisbon ACC100 Card

---

**Dell Technologies, VMware, and Mavenir 5G O-RAN Reference Architecture**

Technical Preview
Reference Architecture Guide
Accelerator card features

The product attributes of the Silicom Lisbon ACC-100 accelerator card include the following:

- Supports eASIC Nextreme-3S family, 35 mm x 35 mm package
- Supports onboard up to 16 Gbit DDR4 with ECC
- PCIe Gen3, X16
- Supports PCI Express Base Specification 3.0 (8 GTs)
- Features a half-length, low-profile design
- Provides a selection of heat-sink types that are optimized to accommodate application power or host server limitations

Network adapter card

Intel Ethernet 800 Series network adapter cards improve application efficiency and network performance with innovative and versatile capabilities. With its high 10/25 GbE port density, the E810-XXVDA4 adapter is designed to optimize high-performance system workloads for cloud and communications.

![Intel® Ethernet 800 Series network adapter](image)

Network adapter card features

The E810-XXVDA4 network adaptor card specifications include:

- Quad Port SFP2PCI Express (PCIe) 4.0 x 8
- PCI Express (PCIe) 4.0 x16
- Application Device Queues (ADQ)
- Dynamic Device Personalization (DDP)
- Support both RDMA iWARP and RoCEV2
- IEEE 1588 Precision Time Protocol (PTP) v1 and v2 support
Conclusion

In this guide, we have provided a high-level reference architecture that is being planned by Dell Technologies and its partners, Mavenir and VMware, for complete testing and release in the second half of 2021. This guide can be used as the basis for future commercial deployments of 5G Open RAN networks. It provides a reusable model to help CSPs accelerate the adoption of this new technology. This Open RAN Solution will provide the advantages needed for next-generation 5G services including:

- A 5G Open RAN solution that provides technology and architectural flexibility to maximize network performance
- Lowered Total Cost of Ownership through automation, and Capex and Opex reductions realized through centralization of key RAN components
- A 5G RAN that moves from a physical to a virtualized architecture, providing simplification of deployment, automation of manual tasks, network resiliency, and ultimately ease of transition to cloud-based architecture
- A RAN platform that will enable increased acceleration of new services to market and that can cost-effectively scale to meet the demand for those services

We believe that this Open RAN reference model, can be a blueprint for CSPs to deploy a 5G Open RAN network that achieves these benefits.

**Note:** This guide may contain language from third-party content that is not under Dell's control and is not consistent with Dell's current guidelines for Dell's content. When this content is updated by the relevant third parties, this guide will be revised accordingly.