ROBIN

# Designing **Open RAN** Platforms

A White Paper

# Table of Contents

## Executive Summary

As the industry moves into the next phases of the 5G rollout, operators face the need to deliver a greater number of new services, with increased speed, lower latency and strict Quality of Service (QoS), at higher speeds and over more endpoints than ever before. This is the lynchpin for delivering higher quality, more profitable services. Among the many advantages, 5G's inherent virtualization and cloud-native frameworks set the stage for not only hardware-software disaggregation, but also the disaggregation of vendors, promoting best-of-breed Open RAN (O-RAN). The bottom line is, if the Network Functions (NFs) that comprise the RAN aren't flexible and high-performing, then none of the valuable Over The Top (OTT) services will be either. Therefore, great care must be taken when choosing NF, cloud platform and orchestration vendors.

The O-RAN framework is being led by a number of providers and vendors that founded the O-RAN Alliance, o-ran.org. The O-RAN Alliance's mission is to “re-shape the RAN industry towards more intelligent, open, virtualized and fully interoperable mobile networks.” O-RAN marks the transition from proprietary to open models that enable operators to select hardware and software from multiple vendors, making them nimbler in building for specific markets. This allows them to engage with a wider array of customers and solutions. Furthermore, O-RAN's open, software-based nature lends itself to automation, and this reduces deployment times and helps operators to build to greater capacity.

At first glance, O-RAN may only seem like a greenfield opportunity. But there are advantages to this approach for those already in rollout or production, by utilizing cloud-native design and automation. In either case, the key component to a successful O-RAN deployment is integration, seamlessly and end-to-end, across protocols and the operations stack. Critical integration tasks include:

- Network Function (NF) performance-tuning on Consumer Off The Shelf (COTS) hardware
- Harmonizing Virtual Machines (VMs) and containers, breaking silos
- Unified, automated lifecycle management of a heterogeneous solution stack, not just NFs and services, driving thousands of clusters and edge data centers at scale
- Protocol and API interoperability

In this document, we will discuss topics related to choosing the right O-RAN platform for your needs.

In all cases, an economical solution calls for flexibility and adaptability as components and requirements will change over time. The foundational pillars for this flexibility revolve around an easy-to-use, intelligent cloud-native platform and customizable bare-metal to services orchestration. Robin products enable successful and scalable O-RAN deployments by providing automated lifecycle management for bare metal, a unified container-based and Virtual Machine (VM) cloud platform for Network Functions (NFs) and Network Services (NS), with automated workflows, monitoring and Methods Of Procedures (MOPs) that support any 3rd party Physical Network Functions (PNF) and applications.

The full impact of 5G applications is yet to be realized and is in its infancy. The right platform and orchestration tools help the operator explore a more competitive and vibrant RAN supplier ecosystem, with faster innovation, leading to deeper market penetration and an improved user experience.

# 1. O-RAN

## 1.1. The Opportunity

According to Dell'Oro, hardware and software will near 10% of the total RAN market in the next five years. It will exceed $10 billion in cumulative revenue over that time, not to mention the additional OTT services it enables. Why is the market so bullish on Open RAN or O-RAN? Because it fills a huge competitive gap in the mobile marketplace. The O-RAN initiative promotes independent development of hardware and software for 5G. Furthermore, it provides a host of new capabilities, and that modernizes your infrastructure, enabling new services and growth potential.

O-RAN and Multi-access Edge Compute (MEC) services are highly complementary technologies. In fact, high-performance O-RAN was designed to enable low-latency, high-speed edge applications. Co-locating the two in a multi-tenant environment makes economic sense, since they can share resources more efficiently than when siloed. Furthermore, in most regions, the revenue generated by OTT applications exceeds that of mobile connectivity. However, to realize these advantages, mobile operators must overcome challenges associated with colocation and security and scale out to maximize their returns.

## 1.2. O-RAN Benefits

New mobile operators are everywhere. The time to create competitive advantage by embracing O-RAN is now. The first step in network modernization is understanding your future demands, choosing paths to revenue growth and then finding platforms that enable them. Intelligent network operators with nimble solutions will seize the opportunity to evolve and lead 5G open solutions.

Key O-RAN benefits enable you to:

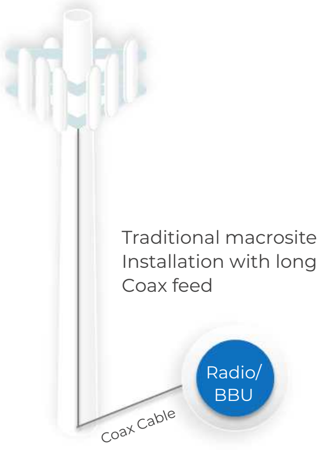
- **Be Free:** The open architecture removes vendor lock-in and stimulates competition. Not only does it open up the vendor market, but it also opens up the integration market and provides opportunities for new specialists.
- **Be Innovative:** Until O-RAN, there was little to no innovation. Also, there was little incentive for vendors to customize to customers or markets. At the very least, O-RAN will open deployment flexibility and accelerate the growth of new OTT 5G services.
- **Be Massive:** O-RAN utilizes automated, hyper-scalable, cloud-native architectures. The scalability and density needed for 5G RAN are tremendous, much more than those of previous mobile generations – more user devices, more traffic, newer services and system requirements. Key cloud technologies, including Kubernetes, that have already revolutionized the cloud, are poised to do the same for 5G RAN and will reduce deployment times, automate lifecycles and reduce human error.
- **Be Nimble:** An open, cloud-native solution dovetails seamlessly into modern DevOps and Continuous Integration and Continuous Deployment (CI/CD) systems, drastically reducing testing and deployment times, enabling providers and developers to iterate quicker. All of this increases business agility and the ability to roll out new services faster, helping you adapt to new revenue opportunities.
- **Be More:** Cloud-native O-RAN can integrate seamlessly into your network modernization plans for edge and MEC applications. With O-RAN, operators will have the opportunity to take full advantage of shared resource pools, growing both RAN and OTT services on demand.

## 1.3. RAN Taxonomy

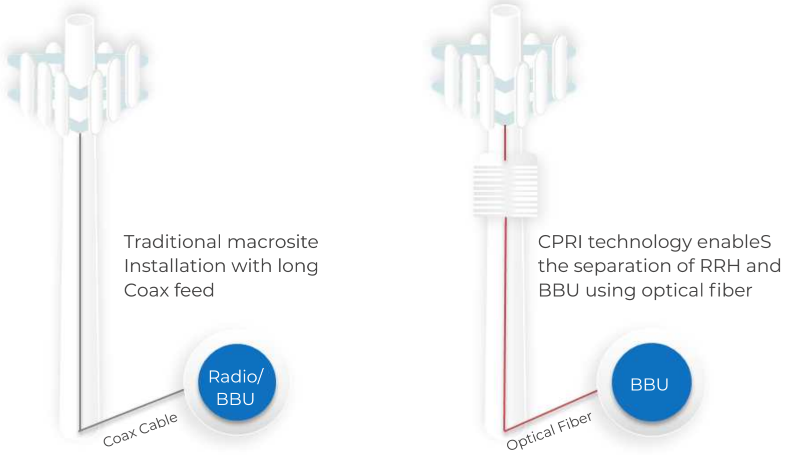
There are a variety of open coalitions and advocacy groups focused on Open RAN. Here, we will take a brief look at the Open RAN industry groups.

**Open RAN**
"Open RAN" will be the umbrella term we will use that represents all open infrastructures that allow one to mix and match RAN vendors.

**Legacy RAN Components**
Originally it was very simple. There was an antenna, that attached via coaxial cable to a Base Station (BTS) device that performed both Radio Frequency (RF) communication and RF signal processing. BTS traffic was then forwarded via a terrestrial transport (backhaul) to the network's mobile core, where call set-up was established and later the traffic is forwarded. All of this equipment was proprietary and part of a closed solution.

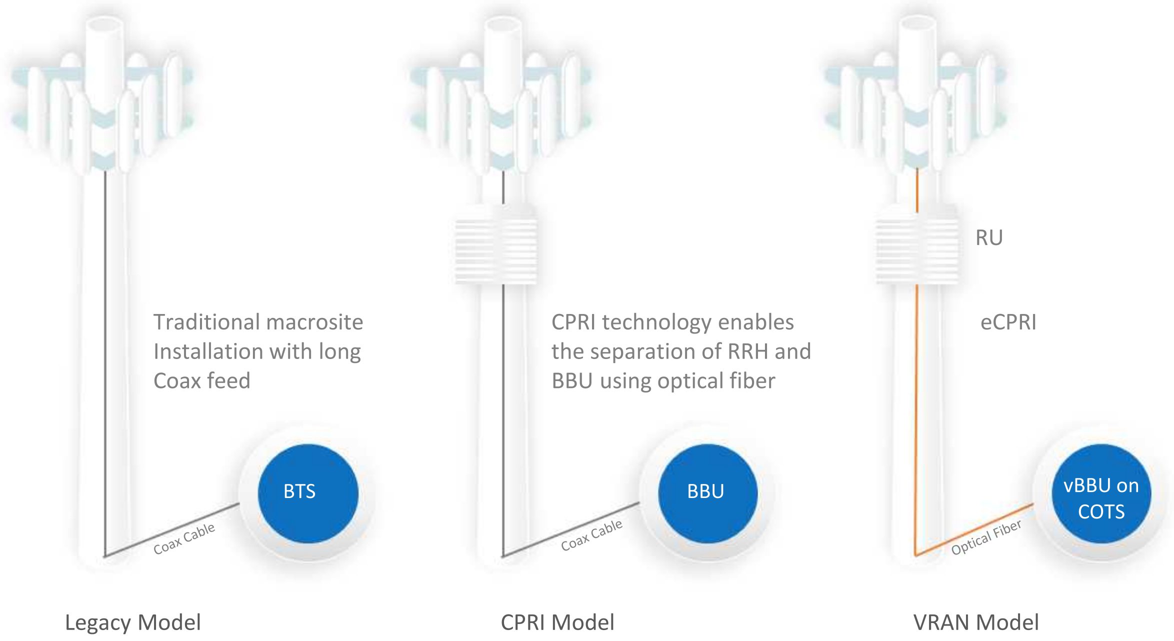


**Centralized/Cloud RAN (C-RAN)**
C-RAN came to be about 10 years ago and began the disaggregation of RAN components. In C-RAN, the BTS was decomposed into two separate devices, the Remote Radio Head (RRH) for RF termination and Base Band Unit (BBU) for signal processing. They were connected via optical fiber using the Common Public Radio Interface (CPRI) protocol. C-RAN was only deployable in areas with access to optical fiber. Unlike its predecessor, the BTS, the BBU could be miles away from the RRH and connected by an economical transport network. Pools of BBUs could be collocated at a central site to accommodate a large number of geographically diverse RRHs. Furthermore, multiple BBUs could be serviced or upgraded at a small number of centralized sites, reducing truck rolls and refresh times. C-RAN was not open, but it did disaggregate RAN components.



**Virtual RAN (vRAN)**

Next came vRAN, which sometimes gets confused with Open RAN. While it is not the same, it did bring virtualization to the forefront. In vRAN, the BBU evolved from a purpose-built appliance to a piece of software running on Consumer Off The Shelf Server (COTS) hardware. Original vRAN software was typically run as Virtual Machines (VMs) on hypervisor platforms. These virtualization platforms could be either proprietary or open. Unfortunately, the proprietary interfaces between the RRH, now called the Radio Unit (RU) and virtual BBUs (vBBU) running on COTS, remained as a barrier to openness. Therefore, RAN networks were still closed silos and single-vendor-driven. However, hardware-software disaggregation was pushed closer to the network's edge. Therefore, there were still many automation benefits to be gained by moving the vBBUs to COTS.


Traditional macrosite Installation with long Coax feed
BTS
Coax Cable
Legacy Model
CPRI technology enables the separation of RRH and BBU using optical fiber
BBU
Coax Cable
CPRI Model
RU
eCPRI
vBBU on COTS
Optical Fiber
VRAN Model

**Open RAN And Its Organizations**

When you use your favorite search engine looking for Open RAN, you will see many different acronyms and hashtags. In many cases they refer to specific groups and specifications in the world of Open RAN.

***OpenRAN Telecom Infrastructure Project***

“OpenRAN” (no space) refers to a movement from the Telecom Infrastructure Project (TIP). The TIP was formed by Facebook in 2016 as a collaborative methodology for building and deploying telecommunications infrastructure. OpenRAN was open to all vendors on vendor-neutral hardware, with software defined, virtualized, technology.

***O-RAN Alliance***

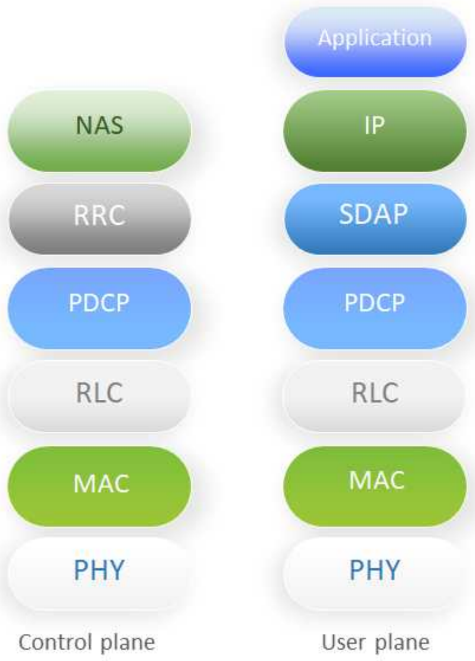
The predominant group leading O-RAN today is the O-RAN Alliance, which was founded in 2018 with the intent on developing standards promoting open and intelligent RAN solutions. It was formed by a merger of two different organizations, namely the C-RAN Alliance, consisting of mostly Chinese providers and vendors, and the XRAN Forum consisting of US, European, Japanese and South Korean operators and vendors. The O-RAN alliance consists of over 30 operators and 200 vendors and has defined a framework and specifications that promote an open, virtual and vendor disaggregated RAN solution.

## 1.4. O-RAN Components

O-RAN adds several new components to the architecture. Some of these protocols will be familiar as they are also used in the Long Term Evolution (LTE) framework.

**O-RAN Protocol Stacks**

O-RAN takes 5G's gNodeB radio protocols and breaks it down into several constituent parts. After doing this, specific parts of the protocol stack are then tied to the NFs that make up the overall O-RAN architecture. O-RAN, as does the rest of 5G, uses Control Plane User Plane (CUPS) separation. Thus, there are two communications stacks, one for the Control Plane and one for the User Plane. The two stacks share some protocols at the lower layer.



**Control Plane Only Protocols**

In the control plane, the establishment and management of sessions occur at the highest layer called Non-Access Stratum (NAS). The main functions of the NAS is the support of mobility management of the User Equipment (UE) and session management procedures found in the lower protocol levels. The NAS is used to establish connectivity between UEs and packet gateways that link them to the outside world.

Next, the Radio Resource Control (RRC) layer is used to exchange control information for connection establishment, system information broadcasting, radio bearer establishment and control of connection mobility procedures.

**User Plane Only Protocols**

In the user plane, the main focus is around setting up Protocol Data Unit (PDU) sessions, that are the means of connection and transport of all user data in a 5G system. The network and the user equipment exchange data via the Internet Protocol (IP). This makes perfect sense as IP is a commonly used protocol for many modern services on the internet and on private networks.

User data then goes through the Service Data Adaptation Protocol (SDAP) layer. The primary function of LDAP is to map the quality of service (QoS) to specific PDU sessions.

### Common Control And User Plane Transport Protocols

For both the control and data planes, the Packet Data Convergence Protocol (PDCP) layer is responsible for data transfer as well as header compression and decompression of IP data, cyphering and data transfer.

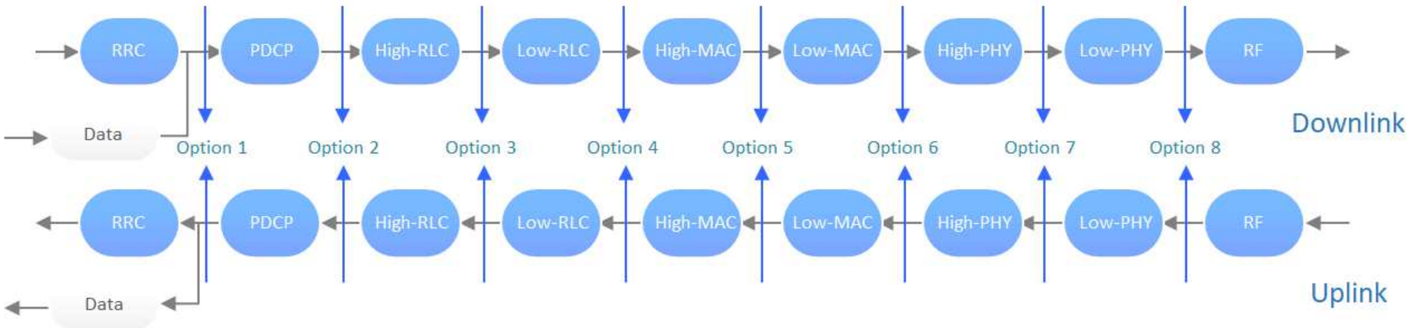
The Radio Link Control (RLC) layer reliably transmits packets over the PDU sessions, for the upper layers. It is concerned with PDU ordering, segmentation and reassembly, deletion of duplicates and retransmission.

The media access control (MAC) layer controls the hardware responsible for interaction with the radio medium. It performs resource allocation and data transfer services to the upper layers and performs actions such as scheduling requests and buffer status.

The Physical (PHY) defines the communication channel to the core network as well as other requirements such as RF modulation and radio beamforming.

### Functional Splits

To muddy the waters a bit, the PHY, MAC and RLC protocol layers are each split into two categories, Upper and Lower and then distributed between the different NFs. Given the functional splits there can be 8 different options.



The choice of how or why vendors/ operators split their functions depends on number of factors that include:

**Increasing performance and availability**

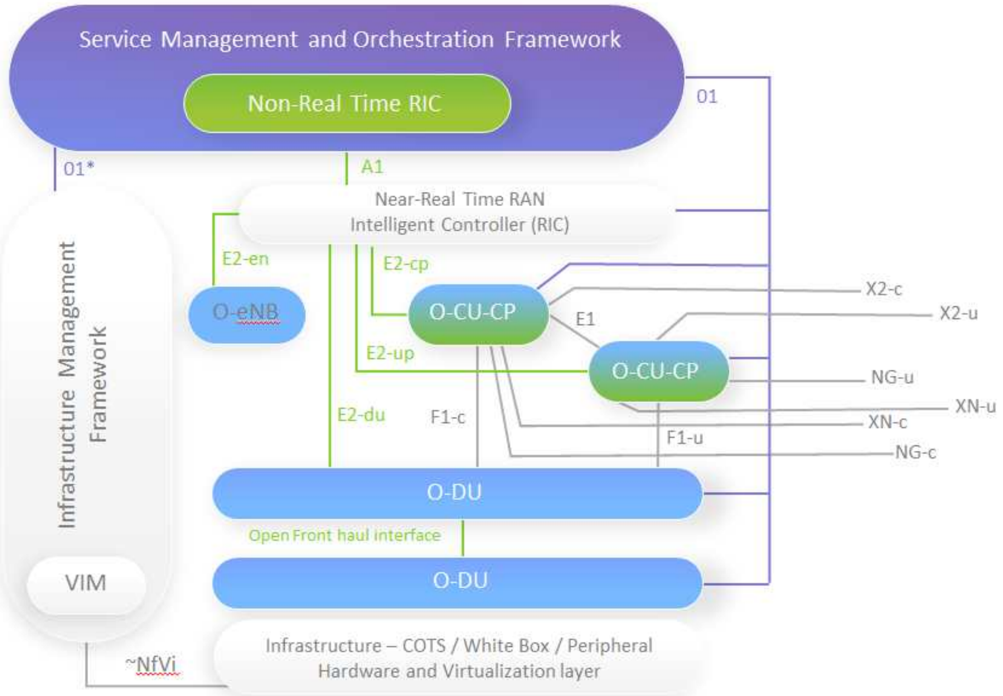
**Performing QoS for specific services**

**Adding additional user density for a given location**

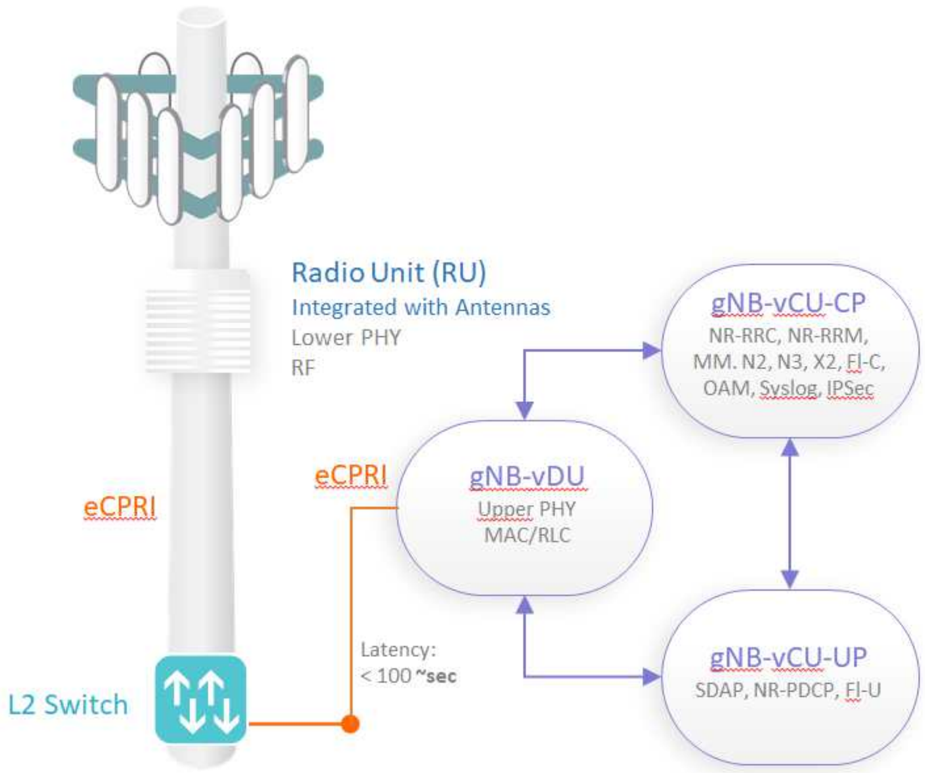
For example, Option 3 is a particular implementation of a split that can occur in the RLC sublayer. RLC functions can be split so that retransmissions may be performed at the High RLC sublayer residing in the Central Unit (CU), while the segmentation may be performed at the Low RLC sublayer residing in the Distributed Unit (DU). This split removes some of the RLC burden from the DU that is placed closer towards the network's edge, where COTS resources may be less available. It pushes it to a CU that is likely located in a site with greater COTS resources. Both the CU and DU are discussed in the following section.

### O-RAN Network Functions

The O-RAN Alliance has several specifications on the makeup of different components and how they interact. The figure below shows how the components and interfaces come together.


Service Management and Orchestration Framework
Non-Real Time RIC
01
01*
A1
Near-Real Time RAN Intelligent Controller (RIC)
Infrastructure Management Framework
E2-en
E2-cp
O-eNB
O-CU-CP
E1
X2-c
X2-u
E2-up
O-CU-CP
NG-u
XN-u
E2-du
F1-c
XN-c
F1-u
NG-c
O-DU
Open Front haul interface
VIM
O-DU
~NfVi
Infrastructure – COTS / White Box / Peripheral Hardware and Virtualization layer

In O-RAN, we keep the radio unit, but the BBU is no longer part of the solution and is split into two different nodes, the CU and DU.



### Radio Unit, RU, also known as O-RU

The Radio Unit (RU) is the digital front-end and includes parts of the physical (PHY) layer, known as the Lower PHY. It also includes other RF characteristics such as digital beamforming. The 5G RU itself is not virtualized and is still a custom appliance. However, unlike previous generations, with O-RAN, its downstream interface is standardized.

The connection to/ from the RU is made over "evolved CPRI" (eCPRI). The term "Fronthaul" is used to label the connection between the RU and the next NF on the journey to the core, the DU. This interface will make use of standard Ethernet frames and User Datagram Protocol (UDP) packets. It can be attached to an Ethernet switch that may be virtualized, physical or integrated into other components.

### Distribution Unit, DU, also known as O-DU

As discussed earlier, BBU is no longer part of the solution and is split into two different nodes. The first of this is the Distribution Unit or DU. The DU can be located at the site of the RU or at an aggregation location. The DU includes physical layer functions known as the Upper PHY and Layer RLC. The DU will run as software on COTS, as will the remainder of the architecture.

### Centralized Unit, CU, also known as O-CU

The Centralized Unit (CU) runs the Packet Data Convergence Control Protocol (PDCP) and Radio Resource Control (RRC) layers. It exchanges control information with the device to set important parameters for the session. The CU runs as software on COTS.  It can be used to aggregate and control multiple DUs. The connection between the CU and DU is known as "Midhaul." The remaining connection from the DU to the 5G core is known as "Backhaul."

The CU can be broken down into two distinct parts, providing Control and User Plane Separation (CUPS) into the CU-CP (Control Plane) and CU-UP (User Plane). While the specification defines them separately, they can be deployed as the same piece of software as long as they adhere to O-RAN interface specifications.

### RAN Intelligent Controller (RIC), O-RIC

The RAN Intelligent Controller (RIC) adds programmability to the RAN network, for added optimization capabilities. It is broken into two parts; however, some vendors may include both parts in a single package or product. The RIC concept is a key deployment accelerator, as it can automate performance. Without the RIC's self-configuring, self-tuning capabilities, radio performance and rollout would become unmanageable.  This automation is critical, given the configuration complexity taken on due to 5G's cell and antenna densification with Massive Multiple-Input and Multiple-Output (MIMO).

### Non-Real-Time RAN Intelligent Controller (Non-RT RIC)

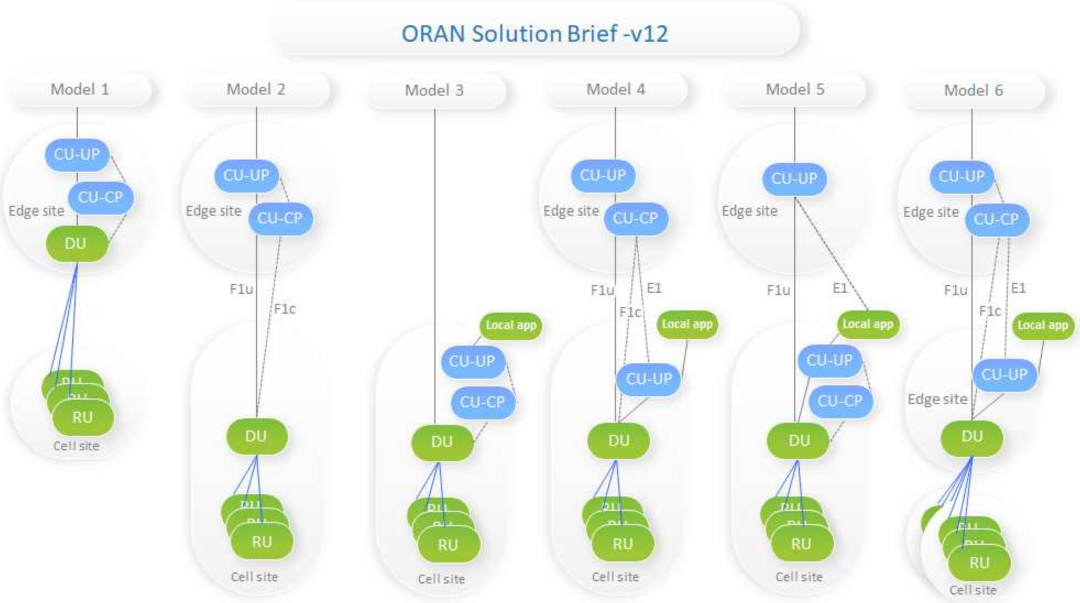
Non-Real-Time RAN Intelligent Controller (Non-RT RIC) performs auto-configuration to all RUs in a plug-and-play manner. Additional functionality includes lifecycle management, configuration, fault-recovery, device management and performance management for all O-RAN NFs. The Non-RT RIC will also support intelligent RAN optimization by providing policy-based guidance, model management and enrichment information to the Near- Real-Time RAN Intelligent Controller (Near-RT RIC).

### Near Real-Time RAN Intelligent Controller (Near-RT RIC)

The Near-RT RIC is used to optimize RAN performance. It may monitor, suspend/ stop, override or control the node via Non-RT-RIC-enabled policies. The Near-RT RIC hosts applications called xApps that run on the Near-RT RIC platform to aid and influence its decision-making functions.

## 1.5. RU, CU and DU Network Design Splits

Given the NFs' highly disaggregated nature, there are numerous networking splits that can occur on a locational basis. Different locational splits have different advantages. In many cases, these splits will be determined by the available COTS resources at the edge and cell sites.



There is neither a one-size-fits-all design nor a cookie cutter pros and cons list as to where splits should occur, as different splits have different advantages.

- The operator may wish to use Model 1 over Model 2 to consolidate resources, but the RU-DU connection will be very delay-sensitive.
- Model 3 pushes all of the NFs to the cell for economy of resources, but this requires larger far cell sites that pose their own challenges.
- Models 4 and 5 allows local MEC applications to be broken out at the edge for a lower latency data service.
- Model 6 combines advantages and weaknesses from Models 1, 4 and 5.

## 2. O-RAN Challenges

**Interoperability – End-to-end Protocols and Operations Stack**

When one thinks of open networks, the first thing that comes to mind is: "How will I make all of these vendors work together?" While that is essential, generally, it is not a stumbling block, as vendors who work in an open environment, centered around disaggregation, usually adhere to standards. They realize that their ability to quickly demonstrate interoperability in proof of concept (POC) is key to their success. Additionally, in today's world of containerized software, vendors can quickly change, test and deploy new code.

Interoperability is more than getting your RAN vendors to properly connect. One must also aim for both end-to-end solution interoperability as well as efficiency up and down the operations stack:

- End-To-End Services (Protocol Interoperability)
    - O-RAN to 5G core
    - UE mobility management
    - UE handover in home and visited networks
    - Inter-operator security points, for example the Security Edge Protection Proxy (SEPP)
- Operations Stacks
    - Harmonizing VMs And Containers to share resource pools
    - Remote bare metal orchestration
    - Remote-tuning of NFs and supporting applications
    - Overlay/Underlay Networking
    - Fast, simplified and flexible instantiation of NFs, applications and services
    - Ongoing lifecycle management

**Virtual Machines And Containers?**

The industry is actively transitioning from VMs to a cloud-native, container-based design and will be doing so for many years. Most legacy cloud platforms support either VM-based Virtual Network Functions (VNFs) or container-based ones (CNFs). Even those that claim to support both typically do so with two separate platforms hidden under a Graphical User Interface (GUI). While this does make for a flashy POC demo, operations, integration, troubleshooting and planning siloes still exist and will hinder a provider's cloud-native journey.

Having one platform for CNFs and another for VNFs is not a migration strategy, it's a technology anchor created from poor decisions. Unaccommodating legacy platforms have the following consequences:

- **Silos limit your deployment flexibility and timelines:**
  All of your vendors have different roadmaps for containerization. When they delay or don't containerize, they force you to deploy more legacy NFs and platforms that make it near-impossible to migrate to a modern, efficient Kubernetes infrastructure
- **Siloed resources reduce utilization:**
  They also impact the number of applications you can run at a given data center, as legacy platforms cannot efficiently share resources among VNFs and CNFs.
- **Siloed operations duplicate operational procedures:**
  Management systems, orchestration platforms, automation schemes, resource pools and training also are duplicated.

**The best step forward is to have a unified platform that supports both CNFs and VNFs on bare metal.**

**NF Performance And Network Tuning**

Your network must scale to support the millions of connected devices predicted for 5G. NF performance will be one of the most important aspects to your 5G success. Low-latency user data transfer is a prerequisite for the category of 5G services known as "Mission Critical Services." These include self-driving cars, emergency response support, remote medicine and drones. These devices cannot function safely in an environment without predictable low latency and jitter. Latency can even apply to "Massive Internet of Things" services, where machines in factories and warehouses interact with people. To achieve high throughput and low latency, every NF vendor has specific hardware and software requirements for the configuration and tuning of their software to run as advertised.

There are numerous categories to classify workloads. They vary in their need for signaling/control plane and user/data plane resources. As one moves from the far edge inward, many of these workloads will share facilities with RAN NFs, some as NFs and some as MEC applications.

- Signaling/control NFs are I/O-intensive with very small packet sizes, which require significant compute resources.
- User/data plane functions are network-I/O-intensive with very high bandwidth requirements, making them both compute- and network-intensive.
- Functions with databases used for data, such as subscriber information, are compute- and storage-intensive.
- Deep Packet Inspection applications, such as those connected to the internet, are highly resource-demanding functions on all fronts.

High-performing NF, MEC applications and the cloud platforms that support them need numerous remote-tuning capabilities to automate and remotely manage performance-tuning parameters, such as:

- Bare metal configuration alteration and Operating System (OS) upgrades
- Firmware updates
- Non-Uniform Memory Access (NUMA) aware resource configuration between cores, memory and networking resources" With "Non-Uniform Memory Access (NUMA) aware resource configuration between cores, memory and networking resources. NUMA -aware granularity is both per-server and per-NUMA node, not just per-K8s Worker Node.
- Complex networking with multiple affinities, IP address, overlays and underlay combinations that connect NFs and extend deep into the operator's provisioning network

- Storage configuration
- Command line parameters executed at runtime

All of these tuning options are required for proper NF functionality and must be remotely manageable at scale.

**At Scale, Bare Metal Provisioning - That Also Enables Performance Tuning**

Given the challenge of performance-tuning and supporting a combination of VNFs and CNFs, it becomes obvious that one needs a very flexible way to remotely provision the bare-metal infrastructure that RAN NFs run on. This becomes increasingly important as many edge data centers are not updated as frequently. In other words, they will likely need to be updated before new platforms, NFs and applications can be remotely installed.

Furthermore, if one hosts third-parties in the edge data center, which is to be expected in 5G, there will be additional MEC applications co-located on the same infrastructure. These too will need performance-tuning, not to mention multi-tenant isolation and role-based access. To roll this out at scale, all of the tuning elements mentioned earlier must be accessible and modellable with repeatable configuration. Otherwise, every edge and far edge data center will become a snowflake and in the worst case a rats' nest.

Regardless, your bare-metal provisioning system is expected to eliminate truck rolls. Therefore, low-level configurations, like OS, Basic Input Output System (BIOS), firmware and Field Programmable Application Integrated Circuits (FPGA), must be manageable at-scale, from far away and over hundreds of thousands of nodes. As pointed out in the previous section, this also includes high-performance networking and storage configurations.

**End-to-End Orchestration, Automation And Monitoring**

Mobile network operators have always had difficulty with rapid changes in demand. Changes in user density or service consumption can overwhelm parts of the network that traditionally saw little use. They can also leave overbuilt areas underutilized. Given that the majority of equipment is moving to software running on COTS, operators positioned to automatically and rapidly adapt to shifts in both RAN and OTT resources will win the day.

To achieve this high level of automation, data center modeling is critical. You cannot be expected to manually configure every Virtual Function (VF) port on every Single Root Input Output Virtualization (SR-IOV) underlay for every Network Interface Card (NIC) that connects to NUMA cores with memory affinity etc. You will literally be programming around the clock and the amount of human error will be disastrous. That is before you even consider redundancy, backup and recovery configs. Your cloud platform and orchestration tools need to take this hunting and configuration out of your hands, by allowing you to model resources for your NFs, applications, networks, storage and data centers as a whole. Additionally, automatic workload placement is critical for at-large-scale deployments as they deploy stop, start, heal and migrate.

Every location, edge, far edge and core can be seen as its own cluster or data center in and of itself. In terms of automation and workload placement, your solution needs insight into physical resources, cloud platforms, NFs and services. In turn, this will enable it to correlate and display views across any strata from a full drilldown to a solution-wide view, including NF, application, service, pod, node, cluster, server, data center, multi-cloud. Based on this multi-level awareness, the information can be used to troubleshoot, auto-repair, migrate, notify and trigger responses using a policy engine.

Another, sometimes overlooked aspect is the ability to correlate and understand what happens in the “what if?” scenarios. Your system needs the ability to explore, plan and automate for failure scenarios. Use blast radius analyses to model and plan for CPU core failure, unreachable network destinations, cluster failures and the like. You need the capability to understand how your network will react when your Key Performance Indicators (KPIs) are not being met.

### Brownfield Deployments

Operators need to begin the journey to better architecture or risk being overtaken by nimble competitors. They must improve existing environments up to levels of performance that can rival greenfield capabilities, without ripping and replacing existing assets.

In a brownfield situation, you still have to overcome the challenges we have outlined to enable successful solutions. The question is: how do you get there, without customer disruption?

Cloud-native technologies are valuable not only for 5G deployments but also for legacy mobility generations. As you modernize your 5G solution, it must accommodate both the old and the new. While on the Open RAN journey, plan to disaggregate the RAN in stages instead of taking a big bang approach, working with vendors to transition in smaller steps. You can define stages based on:

- New locations or scheduled location upgrades
- New service deployments that require an updated RAN
- New network slice introductions
- Addition of new MEC applications
- Inclusion of new Private 5G networks
- End of Support (EoS) milestones: Even though 3G and 4G solutions are not going away any time soon, the appliances and servers they run on will go EoS. This EoS phase provides a path of lesser resistance for improvements and modification.

At any of these points, the RAN may be composed of VNFs, CNFs, bare-metal applications and appliances. This is the ideal time to introduces new orchestration and cloud platforms that harmonize these models, instead of adding multiple silos. NFs of one type can be migrated as-is to the new platform, and even if they are not updated, to a modernized container paradigm. This will enable you to cap your old infrastructure, grow your new one and replace the old as it ages out.

## 3. Open RAN Calls For A Flexible And Adaptable Solution!

Deploying a regional, national or international 5G network is by no means a simple task. As they look to maximize revenue per deployment and compete against nimble specialists, traditional providers will need additional flexibility, as one-size-fits all and even small-medium-large sizing does not make one competitive.

### The Key Takeaway Is Change

We are still in the beginning of the 5G killer app revolution. What your customers need today will not be what they want tomorrow. Thus, they will make decisions on which operator to use, based on your ability to accommodate that change.

- There is no one-size-fits-all solution.
- Resources, NFs, NF vendors and applications will be deployed and migrated all over the network, far edge, edge, regional and core, where each environment has different restrictions and configurations.
- Automation will increasingly influence your ability to quickly rollout new Open RAN components and services at scale.

The key to taking advantage of change effectively is adopting a flexible cloud platform and orchestration toolset that makes it easy to manage your bare-metal infrastructure and move workloads around the network. It also involves reusing resource and network models and existing workflows. Otherwise, you end up re-customizing, re-integrating and reinventing the wheel every time you make a change.

# 4. Kubernetes' Advantages For 5G RAN

Kubernetes is the first Cloud Native Cloud Foundation (CNCF) project and was made public by Google. It is the fastest growing project in the history of Open-Source software, after the Linux operating system. Many see Kubernetes as becoming as ubiquitous as Linux. It is already proving to be indispensable in the cloud.

Why are so many organizations choosing Kubernetes and migrating to containers? There are a number of reasons that center around flexibility, agility and performance.

**Increased Performance**

As we discussed in the "Challenges" section, high-performance-tuned RAN NFs and MEC applications are of utmost importance. Resource performance efficiency becomes increasingly important as solutions scale to the edge and far edge. Here, infrastructure growth, through adding a new equipment rack or building a small facility, becomes both less likely, requires more time implement and is costly.

Kubernetes was designed for performance and scale. Containers' predecessor, virtual machines (VMs), required significant overhead for every application. At the very minimum, this included additional operating systems (OS) called guest OSs, where each guest OS required adaptation from the hypervisor – a software that emulates resources in a highly I/O intensive scheme. Kubernetes completely does away with the guest OS, drastically reducing overhead and OS licensing costs. This means drastically fewer resources, and increased performance and application instances per data center.

**Efficient Scale & Self-Healing**

Automation is a critical factor to the success of any 5G RAN. Autoscaling and auto-healing can greatly improve solution reliability for any application. There are a number of advantages for Kubernetes on this front.

Most Kubernetes applications are broken down into their constituent parts or functions, called micro-services. Let's look at VMs. In order to scale just one part of an application or a simple function, one needs to instantiate an entire VM, including an additional guest OS, and all of the compute/ store/ network resources associated with it, even if it was just to address the need to scale one simple function. With Kubernetes containerized micro-services, one only needs to scale out the micro-service dedicated to a particular function. This can be done in seconds, and only uses a fraction of the resources.

Kubernetes is easily configured to autoscale micro-services based on a number of Key Performance Indicators (KPIs). For example, CPU usage of 80% can be used as a trigger. With similar declarative automation, Kubernetes heals itself when there is a discrepancy between the declared optimal state and any suboptimal state – unreachable, malfunctioning resources or crashed – where each state can trigger a different automated response.

**Open Source & Multi-vendor With Less Risk**

Just like in the NF vendor space, in the cloud platform space too, nobody likes vendor lock-ins. Until recently, for legacy mobility solutions, this was the case. Just as Open RAN pushed an open solution in the radio network, Kubernetes enabled this at a cloud-platform level, making it easier to mix and match vendor 5G RAN NFs and MEC applications. This gives companies the freedom to select smaller, leading edge vendors with far less risk.

Kubernetes is a fully open-source, community-led project overseen by the CNCF. It has several major sponsors – both venders and operators. No one group dictates how the platform develops. To many businesses, this open-source strategy makes Kubernetes preferable to other solutions that are closed and vendor-specific, thus facilitating multi-vendor interoperability without lock-in.

### Portability

There are many regional considerations that impact infrastructure choices, as to where parts of the overall solution are deployed. Edge data centers for O-RAN NFs and MEC applications will likely be heterogeneous, as they may be remotely located and built ad hoc. It is highly unlikely that a particular operator will run applications on the same environment everywhere. Therefore, it is of utmost importance that any software runs effectively in different environments.

Portability indicates whether or not an application can be easily, and in some cases with zero-touch, be adapted to run in different environments. Portability allows you to run services anywhere, instantly, without additional constraints or timely software adaptations that add to a solutions' delivery timelines.

**Kubernetes is highly portable between environments and supports many container runtimes – programs like Docker that run containers. It operates on virtually any type of underlying information technology infrastructure (compute/ store/ network) and does not care if it exists on a private cloud, public cloud or hybrid cloud.**

### Multi-cloud

Similar to portability, multi-cloud allows solution providers and their customers to deploy wherever the resources are. It gives them the opportunity to choose the best locations for those resources.

Kubernetes can host workloads running on a single cloud as well as those spread across multiple clouds, operating in different environments. Kubernetes can easily scale its environment from one cloud to another. This technological agility leads to business agility. With the right Kubernetes platform, one can deploy anywhere at any time, without struggling with migration tasks, platform dependencies and configuration details. Just point it at your cloud and go.

### Increased Developer Productivity

From its early years, Kubernetes was designed to be DevOps friendly, enabling development teams to iterate, test and deploy faster. This is critical in a 5G environment that is continuously innovating and improving on existing designs.

Kubernetes, with its declarative constructs and operations-friendly approach, has changed deployment methodologies. Kubernetes applications use a highly modular approach that enables faster development with smaller, more focused teams that are each responsible for specific tasks. This modularity makes it easier to isolate dependencies and make use of well-defined, well-tuned, reusable and smaller components.

The Kubernetes deployment structure lends itself to controlled rollouts across clusters, canary deployments and automated rollback plans. Teams can scale and deploy multiple times – faster than they ever could in the past.

### Real-World Proven

Kubernetes is a leading cloud solution today. It has been deployed at almost all provider and enterprise markets, with high availability and scale. According to Enlyft, Kubernetes is deployed by over 25,000 companies, irrespective of vertical, revenue and size.

# 5. Kubernetes Challenges

Not all Kubernetes platforms are equal – NFs and provider networks bring new challenges. While Kubernetes is the north star for of edge computing, it was not originally designed for provider solutions. Therefore, with most platforms, one must consider the following challenges:

- Container-CNF and VM-VNF roadmaps and silo implications
- Automated workload placement
- Advanced networking requirements
- NF and application performance tuning
- Declarative automation and orchestration
- Declarative Automation and Orchestration

**CNF/ VNF Roadmaps And Silo Implications**

As discussed earlier, in the "O-RAN Challenges" section, running CNFs and VNFs on separate underlying platforms reduces resource utilization, and adds operational cost and complexity. It also ties a provider's modernization timelines to someone else's roadmaps.

Even if a vendor has a slick GUI linking together "multiple" platforms, operational siloes can still exist. There are still multiple platforms to integrate, install and troubleshoot under the covers. Additional features come at the expense of more licenses and API integration tasks between these platforms. Troubleshooting and complexity doubles, and this poses several questions.

- How does one manage services and service chaining NF lifecycles, network connections and resources across the different VNFs and CNFs?
- Which system does the scaling and auto-healing?
- Can you even service-chain your VNFs and CNFs in the same service – just how difficult is that?

**Robin Cloud Native Platform (CNP) runs both containers and VMs in the same cluster.** They share resources in a common pool, while using the same onboarding procedures. This allows providers to run more applications than before. Existing VMs run 30% faster on Robin CNP than traditional VM platforms such as OpenStack. This is customer-tested and proven.

One of the reasons why customers choose 5G RAN, is for low latency services. Low latency allows applications to connect and make decisions quickly. It is used for mission-critical services, for example factory automation, medicine and autonomous devices. This requirement means deploying services and data centers at the edge. Critical design criteria for edge data centers includes physical space, power and cooling that dictate resource density. At the edge and far edge, this means resource density is at a premium. Robin products enable that value, by reducing overheads with fewer software components and increased NFs and application performance, allowing you to do more with less.

Robin CNP is one system, built from the ground to unify VM and container service operations. With CNP everything, including your CNFs and VNFs, runs on Kubernetes. CNP consolidates all of the resource and lifecycle management operations into a single system, both on the GUI and resource-sharing front.

After you dig deeper into your CNF/VNF vendors roadmaps, you see how their lack of flexibility impacts your service delivery timelines on a modern container platform. Not every application can evolve to containers, especially the older legacy ones. This means you are stuck with multiple design anchors. These legacy applications may be low on your vendor's revenue growth plan, but you still rely on them. At some point you will need to transition them to a software model. If they already run as software, the servers they run on will eventually age out. When this happens, you want to be able to transition them to take advantage of a modern Kubernetes network. But you can't do this if you are stuck on your vendor's timeline. This will severely impact your agility in the world of 5G and MEC applications where only the agile capture the market.

Now for the pertinent question:

- How can you modernize your network in a multi-vendor environment when every application has a different containerization roadmap?

**The answer to this problem is Robin CNP.**
CNP runs both VNFs and CNFs on the same Kubernetes platform, with a unified operations model and fully shared resource pools.

**Automated Workload Placement - Declarative Modeling Vs. Configuration**
Robin's, best-of-breed Kubernetes-based CNP combines **1-click application onboarding** with declarative, context-aware workload placement, pinning your NFs and services to automated policies. This enables **automatic allocation of resources** as they start, stop migrate and heal.Robin CNP makes this plug-and play through Non-Uniform Memory Access (NUMA) aware, auto-resource discovery." With "Robin CNP makes this plug-and play through Non-Uniform Memory Access (NUMA) aware, auto-resource discovery. NUMA-aware granularity is both per-server and per-NUMA node, not just per-K8s Worker Node. Coupled with our unique workload placement algorithms and processes, we can model your needs, then automate and configurate them for you throughout the service lifecycle.

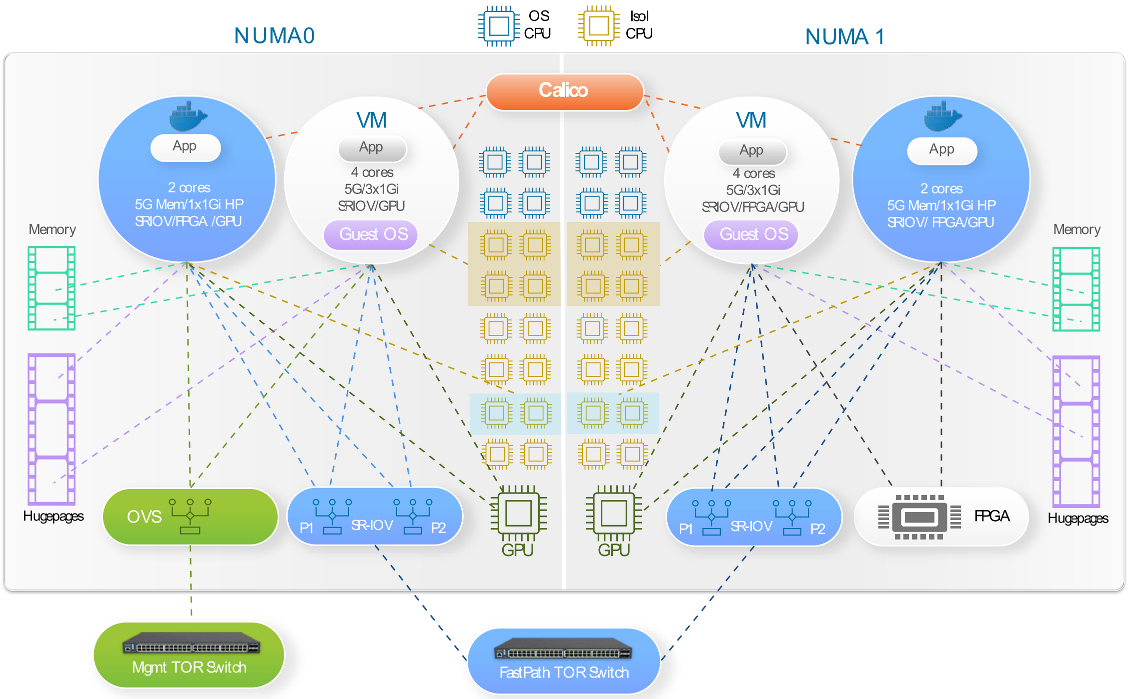
Most legacy platforms force you to manually configure things like CPU cores, NIC cards, physical ports, virtual ports, pods, nodes, VLANs, and pretty much everything network wide. For example, you usually have to explicitly select and configure NUMA node X: CPU1 Core 1 and CPU1 Core 2 on Server Y, and then further configure memory and networking options manually, cross-checking with any other config action that came before it. This is extremely time-consuming, complex and prone to human error. Designing for failover scenarios and enabling automated failover add even more wrinkles.

Robin's user interaction paradigm centers on declarative programmable models, where it prompts you for your desired outcome and not all of the configuration steps to get there. This is possible, because Robin presents all of the configuration parameters as programmable variables. For example, you may wish to reserve a similar configuration from the previous scenario. The request isn't necessarily based on those two specific cores or servers. It is based on a specific "need" - for example an application which requires siblings in the same NUMA node that happen to be Core 1 and Core 2. It could also be based on a different need. But it is still based on some need that can be met by any number of cores in the system, not just those two. So instead of making the user search component by component, Robin interfaces ask you for the need and do the finding and configuration for you. For example, you say "Give me 4 CPUs in the same NUMA node, 1G memory in the same NUMA node, 2 redundant SR-IOV ports and persistent IP addresses", then Robin secures and configures them to support the MEC application. Robin platforms can perform automated workload placement, with awareness that spans not just physical nodes or clusters, but across your entire infrastructure. This eliminates the need to hunt one by one for the right node type for your workload.

As your containers and VMs scale, restart, heal or migrate, those policy declarations are automatically reused to adapt to real-time events. That means your MEC applications are pinned to explicit policies that you only need to set up once.

The diagram below shows 4 NFs modeled for specific resources, all of which are autodetected, based on simple declarative input, without having to know the underlying details and specific resource tags. This entire model can be instantly one-clicked or auto-instantiated into existence on any cluster in any data center.

High Performance NF Data Path



### NF And Application Performance Tuning

In the "O-RAN Challenges" section, we discussed the wide variety of performance needs for the different types of workloads found in RAN and co-located in the same data centers. These can be processor, memory, storage or networking intensive or any combination of the four.

In your multi-vendor O-RAN and MEC environment, each vendor will have different underlying hardware, runtime and command line requirements. They may need to have explicit communication to supporting software databases, message queues and the like. Low-latency user data transfer is a prerequisite for the category of 5G services known as "Mission Critical Services." These devices cannot function effectively in an environment without predictable low latency and jitter. Furthermore, tuned, high-performing software solutions use fewer resources and are more stable over workload shifts.

High-performing NFs, applications and supporting software and their cloud platforms, need the following capabilities to automate and remotely manage performance tuning parameters, such as:

- Basic Input/Output Services (BIOS) versions
- Operating System (OS) versions and configuration
- Field Programable ASICs (FPGA) updates
- Real-time Kernel specifications, for better scheduling control
- NUMA affinity between cores, memory and network underlays
- Isolated sibling processors
- Logically and physically diverse Single Root Input Output Virtualization (SR-IOV) underlay networks for high throughput with low jitter/ latency
- Field Programmable Application Integrated Circuits (FPGA) updates
- HugePages support
- Multiple IP addresses
- Storage network configuration
- Command line parameters used when executing the NF or application
- Automatic Graphical Processing Unit (GPU) model classification
- Robin platforms can perform this tuning, with awareness that spans not just physical nodes, or clusters, but across your entire infrastructure, whithout the need to hunt one by one for the right node type

All of these tuning options are required for proper NF functionality.
**They must be remotely manageable at scale.**

### Advanced Networking Requirements

To Robin CNP, networking is just another resource that is modellable and reusable.

### Network Modeling, Network-as-a-Service

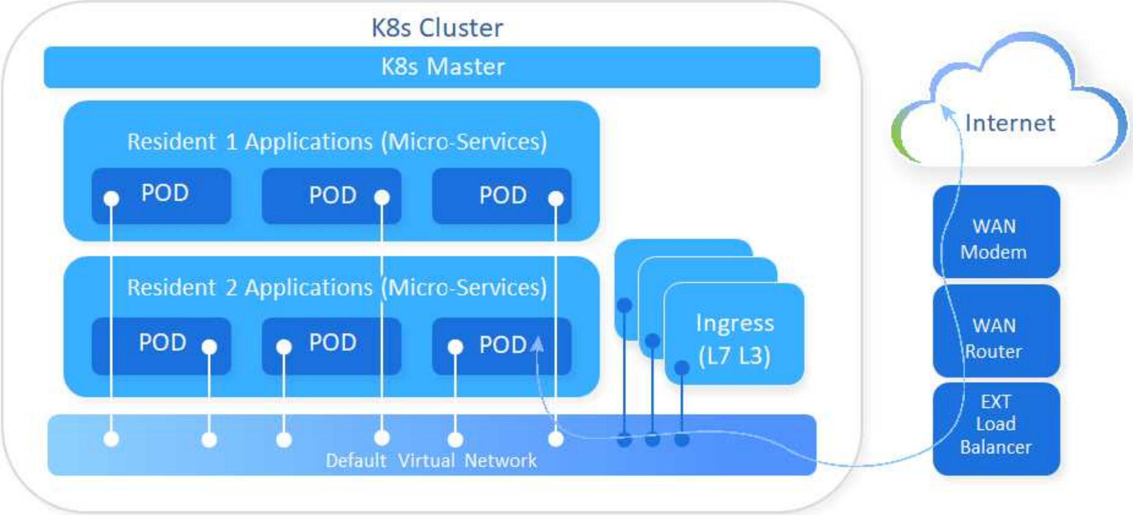
With Robin CNP, you can define multiple network models with connectivity to Calico overlays, connections to Open vSwitch (OVS) underlays (that can be used to interconnect co-dependent NF pods) and SR-IOV underlays (for high-performance traffic). You can even configure Role-Based Access for any underlay or overlay in a network definition. Now you have a set of networks you can instantiate on demand, for any particular user or service request. Adding another cluster? You have a network for that. Moving to a larger or smaller data center? You have a network for that. Setting up a new development site? You have a network for that.

A practical application for Network-as-a-Services is deploying NFs. For example, a Distributed Unit (DU) or a mission-critical MEC application needs networks X, Y, Z, a Calico connection and an SR-IOV interface with two bonded interfaces, over range of IP pools. Defining the network model is that simple, and it takes just one click to deploy and reuse. You can also combine the network models with other elements, including worker nodes. Therefore, you can easily add a new work environment into an existing Kubernetes cluster or even enable it as a self-service option with charge-back capabilities.

As described earlier, regarding DU-CU connectivity O-RAN, there may be several different connectivity models, as well as those of the 5G Core networks and hosting environments. All of these are modellable and available on demand with Robin CNP.

### NF Networking Requirements Are Different From Traditional Cloud Applications

Network connectivity for traditional cloud applications is simple. In most cases, they can connect via a default Kubernetes network, with a simple overlay network separating nodes into separate subnetworks, connecting the nodes, with perhaps an internal ingress load balancer thrown in the mix. Then at some point, the Kubernetes services are attached to an external network with a router, security appliance(s) and external load balancer.

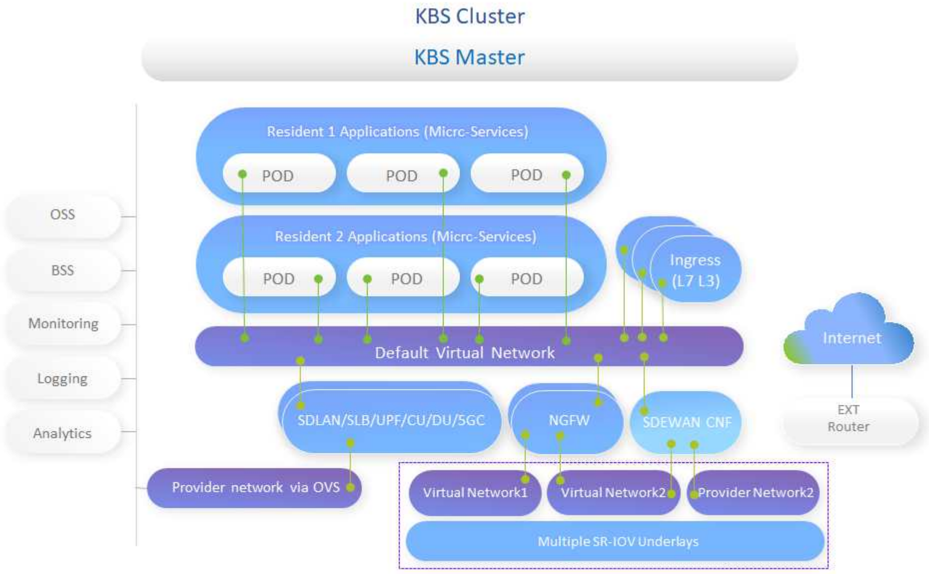


Providers' NFs need more robust connectivity options, that also include high-performance underlays, to deliver the high-throughput, low-jitter services found in 5G applications. These additional networking requirements are not wholly addressed by most cloud platforms and include:

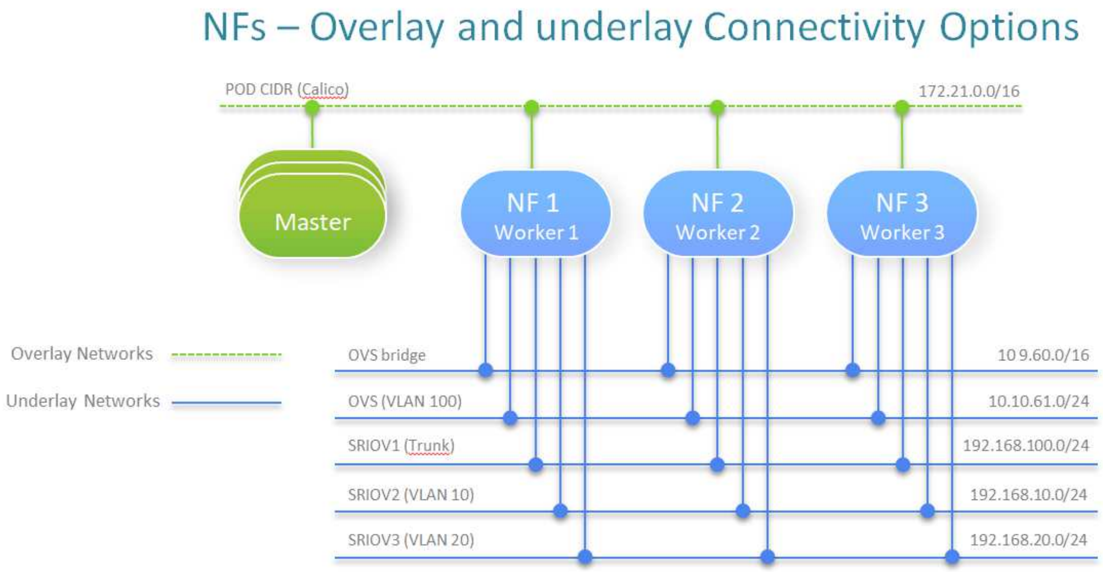
- Per-pod multi-IP network support
- Open vSwitch underlays to extend corporate operations networks to NFs
- SR-IOV underlay networks for high throughput, low jitter, redundancy as well as NF interconnect
- NIC bonding for redundancy and throughput
- IPv4/v6 support
- IP persistency
- Other quality of life enhancements

A typical provider Kubernetes networking design is shown below.

### Functional View



### Under the cover network view



### Per-pod Multi-IP Network Support

The multi-network connectivity shown in the diagram is made possible by Multus.  Multus is an open-source project that enables Kubernetes pods to attach to multiple networks and support multiple gateways, which is table stakes for NF services. It does this by acting as a 'meta' plug-in that calls and intermediates with other Container Network Interfaces (CNI) plugins. It enables the connection to the underlay networks, such as SR-IOV and OVS, as well as overlay networks such as Calico, Flannel, OVN-Kubernetes, Kube-OVN and so on. Thus, the pods have connectivity to both high-performance networks and networks that provide optimal pod-to-pod, or pod-to-outside-world connectivity.

### Layer 2 Connectivity Across Nodes

There are numerous examples for extending Virtual Local Area Networks (VLANs) into Kubernetes. This is a critical component when connecting to legacy operations systems, and can involve Operational Support Systems (OSSs), Business Support Systems (BSSs), monitoring, logging and provisioning tools. For many incumbents, these systems are not available in VMs, let alone containers. There needs to be a simple method to integrate and communicate. The most efficient way is by extending VLAN connectivity directly to the Kubernetes NFs via OVS.

### SR-IOV Underlay Networks For High Throughput, Low Jitter And Redundancy

A critical networking example is extending high-performance VLANs across nodes, for service chains or groups of NFs that expect to be directly connected. Even standalone, mission-critical NF traffic needs to be reliably fast and without jitter. This cannot always be guaranteed with an OVS underlay.

Robin CNP networking improves NF performance by pinning communication to an accelerated path, using SR-IOV underlay networks. But there is more to it than just connecting to the high-performance underlay. We have already mentioned the benefits of auto-configured NUMA efficiencies. Based on user-defined NUMA rules, specific SR-IOV ports sand SR-IOV Virtual Functions (VFs) will be auto-configured. Thus, Robin CNP can guarantee a high-throughput, low-latency data path, with logical and physical diversity. This is extremely important for any low-latency, mission-critical service.

### NIC Bonding

Another CNP modeled element is NIC bonding. We commonly see bonding in SR-IOV underlays. The main benefits of bonding are increased throughput, and load balancing to ensure network continuity in the event of hardware failure.

Robin CNP can auto-select diverse facilities within a NUMA node for maximum performance, or in separate NUMA nodes for maximum diversity. Redundancy can be based on facility-level, port-level or VF-level requirements.

### IPv4/v6 Support

As we discussed earlier, without the right platform, the ability to consolidate and modernize your network can depend on your vendor's roadmap, as we discussed. If your cloud platform does not support IPv4 and IPv6, the platform itself becomes the bottleneck to innovation.

Robin CNP is a dual-stack solution that supports both IPv4 and IPv6 on interfaces and the same pod.

### .IP Persistency

A critical piece of your service reliability puzzle is IP address persistency across the entire lifecycle of a service. Even with lifecycle management, lack of IP persistency across restarts, moves and heals will degrade or completely interrupt service. A traditional Kubernetes IP Address Management (IPAM) plug-in can make IP addressing and subnetting easier for Kubernetes nodes. But it does not enable the persistent IP addresses needed by NFs or any other mission critical application.

Robin's enhanced IPAM plugin enables persistent IP address across all lifecycle events.  Coupled with policy pinning, you can migrate a service from one cluster to another without worrying about IP address or resource implications. These clusters can be anywhere, from regional data enters, to edges or the far edge.

### Additional Quality Of Life Enhancements

#### *metalLB*

As is with all cloud-native services, load-balancing is a critical component to scaling and service reachability. metalLB is a software-defined load balancer with hooks into your Kubernetes cluster. It is lightweight, functional and free. It comes pre-integrated into Robin CNP.

#### *CNI Plugin Customization*

Robin realizes that multi-vendor solutions don't always operate exactly as they are spelled out in the design document. As you build a Kubernetes solution, new requirements also tend to emerge. Robin has customized and rolled many plug-ins into CNP to handle such extremely customer-driven use cases.

#### *Source-Based Routing*

Robin also supports automated source-based routing. This ideal for solutions that have multiple, high-performing underlays between data centers. We typically see this between the edge and far edge. For example, a CU connecting to multiple DU underlays is used to replace multiple static routes that get added as one adds more and more DUs to the network.

## Bare-Metal To Services Orchestration

Robin Multi Cluster Automation Platform (MDCAP) orchestrates and manages the lifecycle of any workflow, including bare-metal provisioning, cloud platform instantiation, applications, NFs, Network Services (NS) and Methods Of Procedures (MOPs), all of which can be auto-triggered through a policy engine. MDCAP's automated workflows support CNFs, VNFs and PNFs simultaneously.

Robin's advantage is that we not only provide intuitive context-aware lifecycle management for your NFs, services and the Kubernetes cloud platforms, but we integrate those workflows with your physical platforms. These include bare-metal servers and third-party appliances. This means one platform, with unifying workflows. Robin MDCAP supports open API, with 100s of APIs to trigger different workflows from northbound OSS platforms, policy engines and Continuous Integration/ Continuous Deployment (CI/CD) systems.

In addition to CNFs and VNFs, there are multiple software-based controllers that aid in the installation of physical devices for PNFs, including RAN radios, surveillance equipment and overlay networks, such as Software Defined Wide Area Network (SD-WAN) controllers, routers and switches. All of this configuration can be kicked off using Robin MDCAP, allowing you to orchestrate the entire deployment step-by-step. Like other Robin workflows, they can be auto-triggered through an interactive policy engine.

Robin makes it easy for providers to deploy new or updated designs down the road. New designs and services can be easily implemented with Robin's extensive workflow engines, where each element in the workflow can be independently modeled, reused and inserted into any existing or new workflow. Robin MDCAP enables you to easily incorporate new things with those that you already know work, instead of reinventing the wheel each time.

### *Building Your Complete Data Center, Starting With Bare-Metal Lifecycle Management Tuned for O-RAN NFs*

As was mentioned earlier, due to O-RANs' need for high throughput and low-latency, the IT hardware platforms need to be highly tuned for every NF and MEC application. Much of this tuning comes in the form of bare metal configuration. Before any application, NF or controller can be installed, a server must be updated, configured and tuned.

Robin MDCAP performs full bare-metal life cycle management and can verify, install, upgrade, configure and bootstrap your server infrastructure. These are not simple scripts, they are **contextually aware workflows** with numerous user-defined checkpoints and forks that guide your installation to its desired conclusion. Transform a server, without configuration or operating system, via the Baseboard Management Controller (BMC), Ethernet or serial, connection.

Monitor and manage readiness:

- Basic Input Output System (BIOS) and BMC configurations
- OS installation, drivers, services and software packages
- NIC, Solid State Drives (SSD), Field Programable ASICs (FPGA), non-Volatile Memory express (NVMe), Redundant Array Of Independent Disks (RAID), firmware upgrades and configuration

Robin MDCAP's bare metal provisioning has been used in numerous provider networks and spans a wide range of pre-integrated Intel and AMD server models.

### Cloud Platform Installation And Configuration

MDCAP is multi-cluster aware and can simultaneously run multiple data centers. Multi-cluster designs are the norm in 5G RAN, since application, NF and control resources can lie anywhere from the far edge to a national data center location.

With MDCAP, you can deploy Robin CNP locally, in the cloud, regionally, at the edge, at the far edge, as a federation as well as in public clouds such as AWS, EKS and Azure. You can use a standardized approach for some cluster designs as well as custom designs for particular locations or services. Like other declarative Robin interfaces, elements in the case of CNP clusters can be modeled and reused. In this case, everything is captured in JavaScript Object Notation (JSON) scripts. MDCAP's editor allows you to open these scripts, modify them, check for errors and enact them all on one motion.

As is with many Robin operations, we define the high-level needs of the cluster by modeling with variables relating to the specific cluster. Then MDCAP and CNP use those variables to configure the desired outcome. Variables include version, High Availability (HA) masters vs. non-HA, OS and firmware versions, FPGA drivers, OS and resource requirements, networks, number of worker nodes and so on.

Once instantiated, predefined automated workflows, such as cluster upgrades, copy, master node changes, moves and deletes are pre-instantiated so you don't have to build them. Customers can also add additional workflows for custom behavior.

### VNF/CNF And Network Services (NS) Lifecycle Management

Robin MDCAP contains several tools for defining applications, NFs and NSs. As was the case of cluster workflows, one can model the key variables, resources, networking, image location and other resource modeling parameters. Then MDCAP provides both prebuilt workflows that cover all of your lifecycle tasks, as well as customization options. Pre-defined workflows include instantiate, start, stop, upgrade, rolling upgrade, terminate and move for VNFs and CNFs.

The only differences between implementing a CNF and VNF are the package definitions. NFs and applications can be onboarded using Helm charts, 3rd party executors, Yet Another Markup Language (YAML) and custom scripts to create Robin bundles that are available in an App store format.

After the packages are defined, there are two straightforward options.

1. Select the cluster you want and one-click deploy any package as a standalone NF.
2. Select the cluster, create an NS chain and deploy. Select the NFs you want, create an ordering and indicate if there are dependencies. For example, don't start NF2 until NF1 is up and running.

When one deploys as an NF or an NS, it doesn't matter if they are VNFs, CNFs or both, since it is all in the package definition. VNFs deployed on Robin CNP perform 30% faster when compared to legacy systems – an important advantage.

Numerous thresholds used for auto-scaling and self-healing can be further defined in CNP.

### Multitenancy and Roles-Based Access

In the case of cloud hosting, shared resources, user self-service and as-a-service applications, MDCAP makes user group segregation easy. MDCAP was built to be multi-tenant. There is no need to run a separate instance for every business entity. For further segregation and deployment flexibility, MDCAP is also configurable for role-based access (RBAC). This can be customized for each tenant. Tenants can not only operate but monitor and explore their solution components in a compartmentalized fashion.

### MOPs Management & 3rd Party PNF Device Support

A MOP is a generic term that describes a step-by-step sequence for performing any task. It tells technicians or automation tools how to execute the actions to perform that task.

MDCAP MOPs facility can be used to manage the lifecycle of physical devices, such as radios, sensors and networking appliances. This is performed by taking existing or new scripts and ingesting them into the MOPs workflow.

MOPs can be executed in batch jobs or triggered by a policy engine that records the change of component states or receives a notification. MDCAP can send values to resident or remote Prometheus collectors that feed the policy engine.

MOPs examples below.

*If an automated process flags that a node, for example an O-RAN radio, goes down:*

1. Notify executives and technician that there is an outage via, email, text and additional alarm.
2. Execute the troubleshooting, repair and traffic migration procedures.
3. Notify customers of an outage.
4. Verify repair.
5. Send the appropriate notifications.

*A technician or a batch job wants to upgrade a Top of Rack (ToR) switch:*

1. Verify there is an alternative network path.
2. Assert the new path on traffic.
3. Upgrade switch.
4. Verify upgrade and configuration.
5. Test connectivity.
6. Roll traffic back to the upgraded switch.

### Monitoring, Automation, Charging and Planning Framework

The key elements to using all of this data are access and correlation. Robin MDCAP and CNP have access to the physical resources as well as logical buckets, based on numerous modeled variables.

Use Robin metrics to gain visibility into the resources across multiple clusters and sites, for troubleshooting automation and planning. See where each service is running, the resources and health status, and collect numerous performance metrics up and down the solution stack, from bare-metal to services.

All monitoring, automation, charging and planning functionalities are available on a per tenant basis.

#### *Monitoring & Automation*

MDCAP and CNP have deep insight into all elements, including physical resources, cloud platform, NFs and services. They can correlate and display views across any strata from a full drill-down to a solution-wide view, including NF, application, service, pod, node, cluster, server, data center and multi-cloud. Using these metrics, one can customize and monitor at multiple levels, not just top-down or bottom-up. Based on this multi-level awareness, the information can be used to troubleshoot, auto-repair, migrate, notify or trigger a MOPs operation using a policy engine. Robin monitoring systems give you true everything-awareness and dependency correlation, as well as the tools to use it.

### Metrics Examples

Assess the configuration, system health, readiness and usage of your bare-metal infrastructure and all the related IT resources. Monitor your Kubernetes clusters. View pod level statistics, cluster health, performance and resource utilization, events, pod relocates, instantiations, terminations, persistent volumes created, volume relocations, disk rebuilds, volume rebuilds, users active, resource pool capacity, node capacity, kubelet daemons, node exporter (exposed via Prometheus), docker daemons, containerd, master status changes and many more.

Get detailed drill-down statistics for services, NFs and 3rd party applications on performance, utilization and health, replica set counts, auto-scale statistics and numerous usage metrics. Monitor and log all MOPs activities.

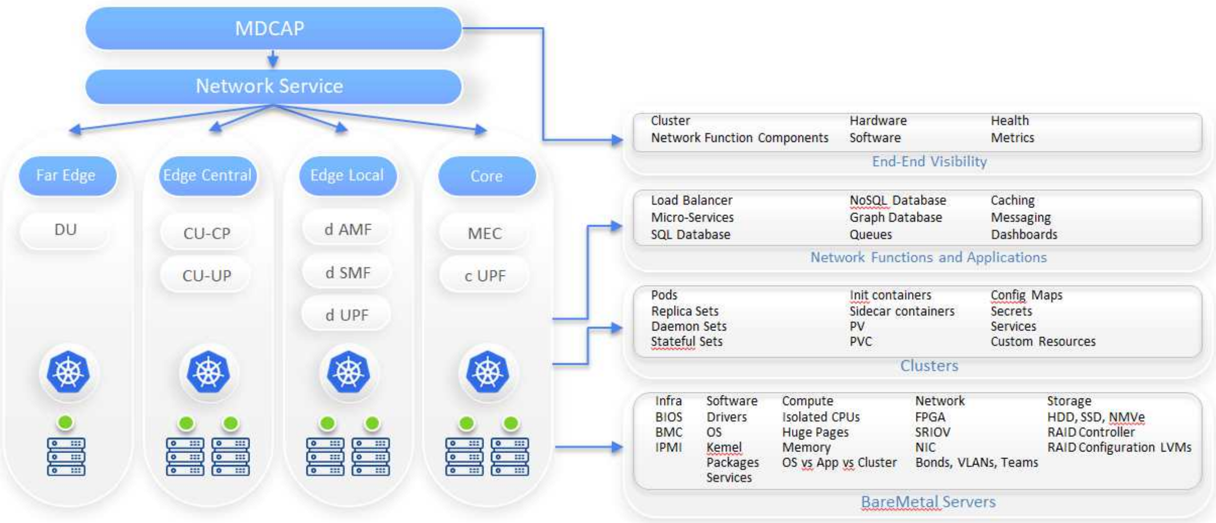
### Charging

MDCAP is capable of exporting all trackable metrics ranging from physical resource utilization, throughput and packet flows, per pod, per node, per cluster. These metrics work in conjunction with our multi-tenancy and role-based access modules for fine granularity among users.

### Blast Radius Planning

Another one of MDCAP's advanced correlation capabilities is the ability to analyze what-if scenarios. Providers can determine the blast radius impact of any system failure including, CPU core failure, unreachable network destinations, worker node restart, master-node failover, roll-out failure, cluster failure, multiple instance failure, or for any other configuration variable, using Robin declarative modeling. They can explore, plan and automate for any failure scenario. Blast radius analyses and automation are not limited to single events but can include multiple events across clusters and data centers.

Inventory Management with Deep Visibility



### Application And Topology Aware Storage

Robin Cloud Native Storage (CNS) supports stateful and stateless applications with industry leading performance. It provides application topology data services, such as snapshots, backup, clones, QoS, replication, encryption, compression, data rebalancing and complex service level management dependencies. When CNS performs a clone, backup, snapshot, replication etc., it not only performs the action on the storage but also stores the:

- Different relationships multiple databases may have in terms of how it is actually used by the application
- The entire application state and configuration
- All associated networking
- Metadata

It also performs pre and post clone operations for one-click migrations that do not require new IP addresses and the like for the new, active, clone.

CNP supports all major Kubernetes distributions: Robin CNP, Anthos, OpenShift, Rancher, AKS, EKS, GKE and open source, to name a few. CNS is included as a part of Robin CNP and is also available independently.

## Robin Conclusion

O-RAN gives operators the opportunity to engineer change in a way that delivers far reaching improvements to operation efficiency, costs and agility, as well as their infrastructure utilization. The operational nature of RAN environments can be transformed to become nimbler, offering greater choice and flexibility to accelerate new connectivity and service options.

Whether it is greenfield or brownfield, there is a simple plan that operators can follow.

- **Identify an insertion point:**
  Choose a specific service, NF, data center, region or service that will benefit from RAN openness.
- **Plan and build for an ecosystem:**
  Find vendors who want to solve "your" problems. Partnering with vendors is important. Look for a win-win scenario where both operator and vendor benefits from a specific innovation.
- **Develop better processes:**
  Take advantage of your new ecosystem that includes cloud-native components and new orchestration methodologies to not just increase the breadth of your services and reduce complexity.
- **Test, tune, scale and insert:**
  After validating interoperability, performance, security and end-to-end service functionality, take your new operations processes and roll them out at scale with a solution that can roll-out, roll-back and migrate without service impacts.

Robin enables you to deliver on the 5G promise with unmatched lifecycle simplicity, performance, scale, and advanced workload placement. By utilizing Robin MDCAP with its bare metal to services orchestration and the industry's most advanced cloud platform, Robin CNP, we enable:

- **Automated and resilient life-cycle management:**
  Full stack life-cycle management of the bare metal HW platform, SW platform, Cloud platform, 3rd party appliances and CNF/VNF services chains
- **Simplified deployment, flexible and high-performing:**
  One-click onboarding, with an easy-to-use declarative model that scales, heals and migrates using autoconfigured, service-pinned policies
- **Industry-leading cloud platform, designed for 5G applications:**
  A highly preferred cloud-native, Kubernetes-based platform that supports flexible networking options, VNFS/CNFs, application-aware storage and advanced multi-parameter, multi-cluster, workload placement
- **Low footprint and high scalability:**
  Scales to over 1,000,000 elements, with advanced features like network-wide monitoring, analytics and closed loop automation

For more information
visit **robin.io**
email **info@robin.io**

224 Airport Pkwy, Suite 600, San Jose CA 95110 USA
Tel: 408.216.0769