AvidThink®

Enabling Successful Edge Services Deployment

The Edge is Not Just Another Cloud

Table of Contents

Wind River Studio Conductor Solution Review

About AvidThink®

AvidThink is an independent research and advisory firm focused on providing cutting edge insights into the latest in infrastructure technologies. Formerly SDxCentral's research group, AvidThink launched as an independent company in October 2018. AvidThink's expertise includes 5G, edge computing, IoT, SD-WAN/SASE, cloud and containers, SDN, NFV, and infrastructure applications for AI/ML and security. AvidThink's clients include Fortune 500 technology infrastructure companies, tier-1 service providers, leading technology startups, prominent management consulting firms and private equity funds. Visit AvidThink at www.avidthink.com.

Enabling Successful Edge Services Deployment

Why the Edge is a Different Type of Cloud

Introduction

Communication service providers (CSPs) increasingly realize that edge computing will play an important role in creating and deploying the next generation of digital-communications-dependent services. Whether powering a virtual or Open radio access network (vRAN, OpenRAN, O-RAN) for a 5G build-out, transforming the automotive experience via cellular-vehicle-to-everything (C-V2X) initiatives, supporting a resurgence of Internet of Things (IoT) and Industrial IoT (IIoT), or enabling new extended reality (XR) services encompassing augmented and virtual reality (AR/VR), new services will need ample and proximate computing and storage resources.
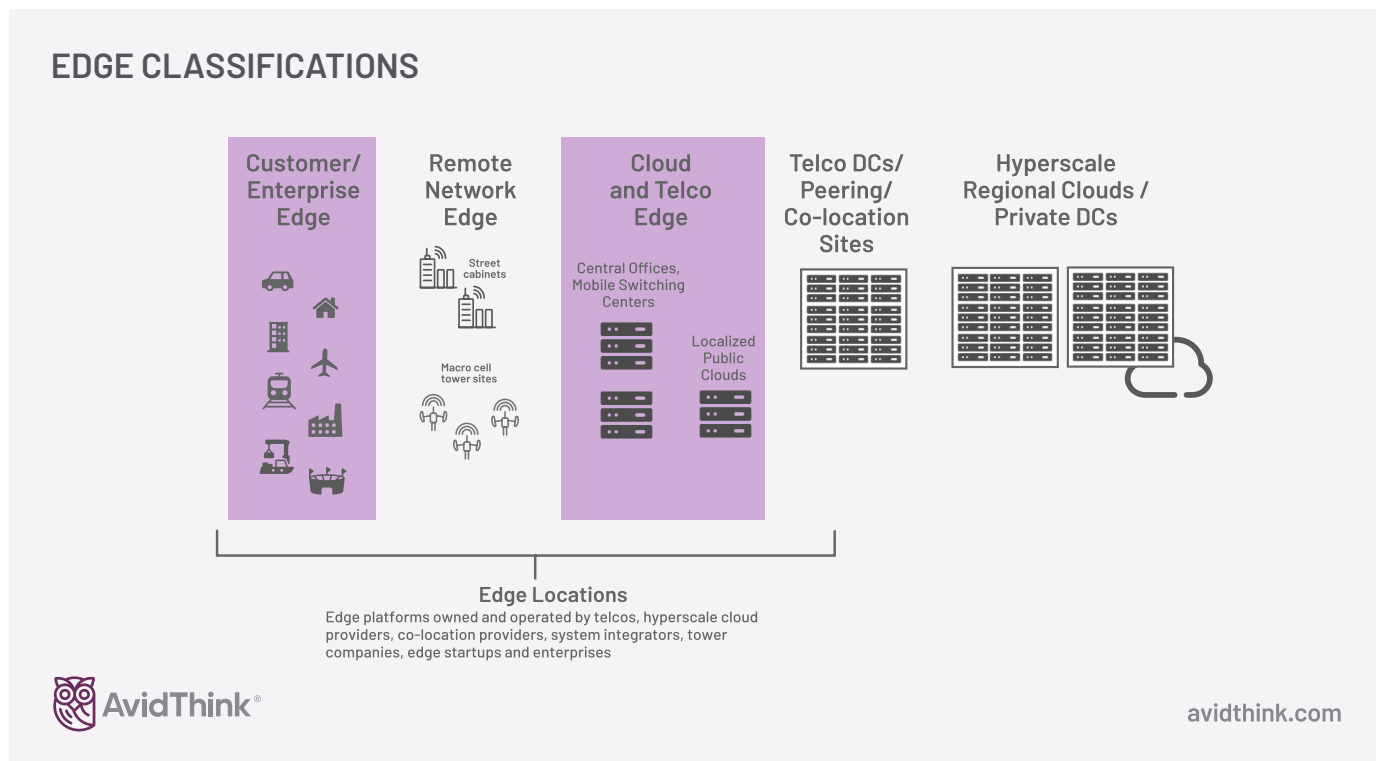
Web 2.0 companies and hyperscalers have honed their techniques, developed new technologies, and established best practices in data center infrastructure build-out. These industry giants have helped improve the scalability, availability, and efficiency of the world's largest cloud data centers. With the expansion of cloud platforms to the edge, web companies, hyperscalers, CSPs, and anyone involved in building out the edge need to understand and accommodate underlying differences between these new sites and regional cloud data centers.

For CSPs, the edge is a new yet critical element in their 5G build-out. The edge is also a component of telco cloud transformation. To successfully leverage the edge, CSPs must understand why and how it differs from today's centralized clouds. This research brief aims to cover these differences in new services and workloads at the edge, including vRAN/O-RAN. We will discuss today's edge challenges, as well as strategies that can improve the CSPs' odds of success in this new endeavor.

CSP-Pertinent Edge Sites

To set the stage for our subsequent discussion, we'll first define the edge locations that CSPs are building out as part of their edge footprint:

- **Network edge** refers to edge computing resources embedded in or close to mobile and wireline networks, situated within the figurative last mile from end-user devices. For convenience, in this research brief, we'll consider edge computing clouds located in major metropolitan areas as part of the network edge. These metro edges may be operated by hyperscalers, CSPs, or both as part of a partnership. There are likely hundreds to thousands of these edge sites per country that a CSP will need to deploy and manage.

- **On-premises or enterprise edge** includes computing resources in private or semiprivate spaces, including factories, event venues, transportation hubs, farms, mines, and oilfields. CSPs with many large enterprise customers, each of which might have hundreds or thousands of locations under management, may be faced with tens or hundreds of thousands of enterprise edge sites.

- **Radio access network (RAN) edge** covers cellular base station-located edge sites, including roadside units or cabinets in dense urban areas. Early indications suggest these sites will be initially used for CSP vRAN/ORAN workloads, with third-party enterprise workloads to follow in the future. For CSPs, the number of RAN edges will number in the tens of thousands.

EDGE CLASSIFICATIONS

Customer/ Enterprise Edge	Remote Network Edge	Cloud and Telco Edge	Telco DCs/ Peering/ Co-location Sites	Hyperscale Regional Clouds / Private DCs
	Street cabinets	Central Offices, Mobile Switching Centers		
	Macro cell tower sites	Localized Public Clouds		

Edge Locations

Edge platforms owned and operated by telcos, hyperscale cloud providers, co-location providers, system integrators, tower companies, edge startups and enterprises

AvidThink®

avidthink.com

New G, New Services

5G is a reality. It's also a work in progress. Early 5G-enhanced mobile broadband (eMBB) services aren't driving increases in consumer average revenue per user (ARPU). 5G fixed wireless access, however, is looking promising as a revenue generator for the mobile network operators (MNOs) and competitive in some markets against wireline digital subscriber line (DSL) and cable technologies.

With 5G in both public and private enterprise settings, the CSPs are looking to bring in new services, or revitalize old services like IoT, to drive revenue. And many of these new services require an edge computing platform.

At a minimum, an edge computing platform is necessary to provide improved 5G quality of experience. Lower latency and improved predictability can be achieved by hosting 5G user plane functions (UPFs) at edge locations. Further, an enterprise edge can host private 5G cores (5GC) on enterprise premises, allowing local traffic breakout and integration with on-premises assets.

In concert with edge computing, 5G can fulfill ultra-reliable low-latency communication (URLLC) service needs. And the edge is a great location to host IoT application modules that preprocess the data: filtering, aggregating, or summarizing before uploading for analytics or storage in regional cloud locations. Likewise, new services that require fast response times in control or safety applications can benefit from hosting intelligence (AI/machine learning) at the edge.

And most critically, the increasing popularity of virtualized RAN and open RAN architectures, particularly those compliant with O-RAN Alliance's specifications,

> **With 5G in both public and private enterprise settings, the CSPs are looking to bring in new services, or revitalize old services like IoT, to drive revenue. And many of these new services require an edge computing platform.**

will push CSPs to build edge infrastructure that can host virtualized components of the RAN. Using standardized server platforms to run virtualized baseband units (BBUs) for vRAN, or virtualized distributed units (vDUs) in concert with virtualized centralized units (vCUs) that connect to a wide range of radio units (RUs) for O-RAN, can achieve improved operational and capital efficiency while bringing CSPs increased flexibility. Many O-RAN workloads are packaged as cloud-native network functions (CNFs) running on containers, so vDUs and some vCUs will be hosted on edge clouds. While SmartNIC and other hardware acceleration might be needed for cost-effective and optimized deployment or to fulfill precise timing and synchronization requirements, O-RAN container edge workloads can be deployed and managed like cloud apps.

Of these workloads, vRAN/O-RAN are likely the most performance-demanding and latency-sensitive today. However, with future workloads coming from the XR space, given the current buzz around the Metaverse concept, understanding how to accommodate these types of applications will be important for CSPs looking to support innovative use cases that represent revenue opportunities.

Edge, More than a New Cloud

From an application development perspective, today's edge software stacks look similar to regional cloud stacks — container and VM workloads can be deployed on edge servers as on regional public or private clouds. However, the cloud abstraction layer hides critical infrastructure differences that both edge infrastructure operators like CSPs and application owners need to understand:

- **Management scale differences:** While regional public clouds and private clouds consist of numerous servers in a few locations, the edge is made up of a small number of servers in hundreds or potentially tens of thousands of locations. This has implications for the management and maintenance of these servers and the software stacks.

- **Small edge footprint with little supporting resources:** Applications running in regional clouds have elastic access to large amounts of storage and computing resources — scaling up instantaneously on demand. In contrast, edge sites have limited resources, and bringing in new equipment takes weeks or months.

- **Real-time and latency sensitivity workloads:** Many public cloud workloads are enterprise and consumer web-based applications powered by large databases and complex middleware implementing sophisticated logic. Edge workloads that help power today's communication networks have greater latency and performance sensitivity than those web applications, which often depend on critical timing to perform effectively. And with upcoming XR applications, putting in place a platform that can accommodate time-sensitive workloads is needed.

- **Extreme lights out, costly truck rolls, no onsite personnel:** Any changes to edge infrastructure will probably require expensive trips to remote locations. Likewise, the types of expertise available at the edge could be limited, with fewer troubleshooting capabilities than in data centers.

- **Lower bandwidth pipes, intermittent connectivity:** The majority of application workloads in the cloud seldom worry about connectivity and available bandwidth. Edge sites, on the other hand, will likely have less network bandwidth capacity and might face intermittent disruptions to the network.

- **Different high availability models:** In cloud application architectures, high availability is usually achieved by putting redundant application components in different availability zones that share little common underlying infrastructure (hardware, network connectivity, energy sources) and are geographically separate. Edge applications will have to determine whether a redeployment to adjacent sites if the primary site fails is feasible or whether components have to be deployed higher up the topological hierarchy, in a regional cloud or metro cloud, with attendant increases in latency.

> **The cloud abstraction layer hides critical infrastructure differences that both edge infrastructure operators like CSPs and application owners need to understand.**

- **Increased hardware diversity:** Public clouds, and many private clouds, often limit the variety of hardware types that they use. This reduces management and inventory complexity, streamlines purchasing processes, simplifies assurance, and reduces overhead. At the edge, due to the different types of workloads, and variety of locations (harsh environments, lack of cooling, limited power), there will likely be a greater variety of hardware that better adapts to these environments.

- **Remote sites with unreliable security:** Centralized data centers are secured by strict physical access, often with biometric protection and constant video surveillance. Remote sites, on the other hand, could be a simple locked shed near a cluster of 5G base stations or a roadside cabinet in a dense urban area. These servers could be stolen, along with the data resident on local storage.

The Need for Edge-Native Cloud Platforms

Cloud management platforms that provide virtualization capabilities and the ability to run both VM and container workloads have served enterprises and cloud providers well in private data centers. Proprietary hyperscaler stacks have been developed that are specialized for large-scale data centers with thousands and tens of thousands of servers — powering centralized infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS), software-as-a-service (SaaS), and serverless workloads.

As we move towards edge-native infrastructure supporting a wide range of CSP workloads, including 5G RAN and core, cloud platforms will need to adapt to become edge-friendly or revamped to be edge-native.

An edge-native platform will retain the workload management capabilities of cloud platforms, including hosting VMs and containers on VMs or bare metal, but will need to address the underlying architectural differences covered earlier.

To address the needs of the edge, key architectural attributes of edge-native cloud platforms must include:

Model-and intent-based: Imposing a strict and consistent abstraction is critical to enabling manageability at scale. Individually managing single servers or even a small cluster of servers is not feasible with thousands or tens of thousands of edge sites. Stating intent and higher-level goals while letting the system autonomously handle the detailed programmatic aspects, target by target, is the only means to achieve the scale the edge needs.

Automation-centric: In parallel with a strong abstraction model, the configuration of edge systems needs to be automated, taking an infrastructure as code (IaC) philosophy where all configuration is codified and managed centrally, disallowing local exceptions and overrides except in emergencies.

Hierarchical structure and group organization: Support for hierarchical structures with the intelligent inheritance of configuration attributes will be needed. For instance, applying certain data management or encryption policies to all edge sites in a country or geographic region or setting the networking policies for all edge sites with hardware-based SmartNICs installed.

Centralized visibility and manageability: As edge systems proliferate, the platform needs to continually provide a single source of truth that aggregates metrics and telemetry across the entire deployment. This centralized pane-of-glass needs to proactively and intelligently highlight potential issues by leveraging machine learning and allow for drill-down as needed to troubleshoot errant edge resources. Likewise, using assisted or auto-remediation to continually check remote systems have not drifted from their configurations is essential to correct operations at scale.
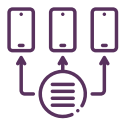
Resilient provisioning: Given the high cost of truck rolls to remote sites and the limited number of backup systems at each site, edge-native systems need bulletproof provisioning. Firmware and software update mechanisms need to be robust, and management systems need to allow for disconnected operations with reliable reset and recovery mechanisms.

Real-time or near real-time performance: With 5G and performance-sensitive telco workloads prevalent on CSP edge sites, the platform will need to expose precision timekeeping and provide real-time or near-real-time assurance and bounds for latency- and timing-sensitive workloads like 5G RAN.

Robust security: Given that there are no administrators or technicians present at remote sites, the edge-native cloud platform must provide high levels of security to support trusted platform and computing systems, ensure that edge systems maintain their integrity, and bolster tamper-resistance signals from the underlying hardware. Likewise, platforms need to support encryption-at-rest and encryption-in-flight through smart key management that destroys data if units are removed from the edge site.

Multi-edge support: The reality is that there will be multiple flavors of edge sites: from private edge to multiple public cloud edge sites. An edge-native platform needs to be able to orchestrate and manage other systems in concert with its primary platform.

My Cloud or Yours - The Edge Multicloud Reality

For CSPs, the edge will involve multiple providers. In addition to CSP-owned private edge clouds, CSPs will likely partner with and utilize hyperscale cloud platforms like AWS, Microsoft Azure, or Google Cloud Platform, or perhaps those offered by Oracle, IBM, or even Equinix Metal.

> **In addition to CSP-owned private edge clouds, CSPs will likely partner with and utilize hyperscale cloud platforms from AWS, Microsoft Azure, or Google Cloud Platform, and perhaps Oracle, IBM, or even Equinix Metal.**

A key driver for this multi-edge reality is the limited capacity at many edges — especially the far edge — coupled with variable workloads that may demand high peak capacity. The huge capital expense and time lag of putting new infrastructure in multiple locations have resulted in a predilection for leveraging already built and available edge capacity. This "edge bursting" will spill over into public cloud edges or other edge platforms which may not be CSP-owned or CSP-operated.

As new business models emerge around neutral host RAN, there's further anticipation that the underlying edge powering that RAN will also follow the neutral host model. This means that CSPs will need the ability to integrate and push workloads onto different cloud and container platforms, some of which might be proprietary. Consequently, the edge-native cloud platform will need to provide plug-in capabilities and be extensible to accommodate new infrastructure platforms that emerge.

Maximizing Success with the Edge

Just as 5G is a reality for CSPs, so is the edge. CSPs can no longer wait to make plans for the edge — especially with 5G rollouts in progress. There are decisions to be made around edge platforms that will overlap with CSP telco cloud decisions.

Edge Cloud Decisions

CSPs that have telco clouds in place will need to determine if their existing telco cloud platforms can be extended to the edge. Walking through the attributes covered earlier in this research brief and determining if they are supported on the existing platform will provide CSPs with an understanding of how edge-friendly their platform will be. In particular, many telco clouds today might not be equipped for the real-time nature of vRAN/O-RAN workloads and should be evaluated for what will likely be a key edge workload, particularly for MNOs.

For CSPs in the process of revamping their telco cloud, or those contemplating an edge-first strategy, there are numerous choices. They could consider edge platforms from the hyperscale cloud providers like AWS Outposts, Azure Stack Edge, or Google Distributed Cloud Edge, and then pair it with AWS EKS and EKS Anywhere, Azure Arc, or Google Anthos for orchestration. Alternatively, they could pick from private cloud platforms and corresponding orchestration suites popular with CSPs such as Red Hat, VMware, and Wind River. Note that these solutions are also primary candidates for telco cloud platforms.

Further, as each of these orchestration suites expands to incorporate support for third-party platforms, CSPs will have to determine whether the cloud and orchestration choices are a joint decision or two separate decisions. If both solutions come from a single vendor, that can reduce finger-pointing and provide increased assurance around the complete stack. The ideal situation would be a single vendor for both, with an underlying cloud platform that is architected to be edge native. CSPs will want to carefully evaluate candidate platforms bearing in mind the guidance laid out earlier in this research brief.

Getting Started with the Edge

Regardless of the CSP strategy for its edge platform or orchestration strategy, success for the edge depends also on other factors. In our work with tier-1 CSPs who have adopted edge computing, we've collected the following observations that we believe can help maximize CSPs' odds of success:

- **Start with an appropriate use case and build the business case** — Instead of starting with the technology platform, or driving immediately to a pervasive public edge, find early use cases that can be deployed in a limited scope. This allows early demonstration of market traction and measurable business outcomes that help build the larger business case.

- **Align edge strategy with 5G RAN deployment** — Find synergies to help improve ROI for edge deployments by amortizing costs as part of RAN development while keeping spare edge resources for external workload monetization. MNOs we work with are using the vRAN/O-RAN workload to drive the requirements for the edge platform, reasoning that starting with the most performance-stringent use case will ensure a foundation that supports all of today's workloads while getting ready for XR and Metaverse-related applications in the near future.

- **Develop a strategy around multi-edge and identify partnerships** — Given the multi-edge reality, start early conversations with hyperscaler, colocation providers, and server OEMs that support the key use cases.

- **Select or build edge orchestration and automation platforms early** — The edge needs to be architected for scale from day one. Manageability cannot be built after the fact. CSPs will want to evaluate the fit of automation platforms available and weigh time-to-market needs versus uniqueness and differentiation in a build-your-own model.

- **Bake in assurance** — Assurance of edge performance, security, and compliance, needs to be part of the deployment of the edge. CSPs will want a deliberate plan for managing telemetry feeds and figuring out the analytics strategy. Without a commitment to co-develop an assurance infrastructure with edge rollout, the sheer scale will overwhelm CSP operation teams. CSPs will want to tie assurance to their automation solutions for a closed-loop management system.

The edge is a reality, and the edge decision is one that CSPs must make today. AvidThink believes that the guidance laid out in this research brief will assist with this strategic yet difficult decision. We're always open to feedback and a conversation. Our team can be reached at **research@avidthink.com**.

Wind River Studio Conductor Solution Review

Cloud-Native and Edge-Native Automation and Orchestration

Introduction

Wind River Systems is a California-based company focused on both cloud and embedded solutions. With an edge-to-cloud software portfolio, Wind River addresses the needs of critical infrastructure companies looking to realize the potential of the Internet of Things (IoT). In January 2022, Aptiv, a global technology company serving the automotive market, announced that they would acquire Wind River from TPG Capital, a private equity firm. The transaction is expected to close mid-2022.

Wind River has built a strong brand in the embedded and IoT space. Additionally, Communications Service Providers (CSP) recognize Wind River's early investment in network functions virtualization (NFV) and recent expansion into 4G and 5G RAN with its ultra-low latency software infrastructure used in many mobile deployments. Current offerings include Wind River Studio, a cloud and edge platform designed for CSPs and industries with critical workloads. Wind River Studio provides a cloud-native platform for the development, deployment, operation, and servicing of mission-critical intelligent systems. Studio enables CSPs to deploy and manage cloud-native workloads, including virtualized RAN and 5G open RAN. The Studio platform aims to simplify CSP operations by providing single-pane-of-glass, zero-touch automated management of thousands (or more) edge nodes.

What is Wind River Studio Conductor?

Wind River Studio Conductor is an orchestration and automation solution. With Wind River Studio Cloud Platform, based on the open-source StarlingX project (hosted by the Open Infrastructure Foundation), and Wind River Studio Analytics, provider of monitoring, analytics and closed-loop automation, they comprise the Wind River Studio suite.

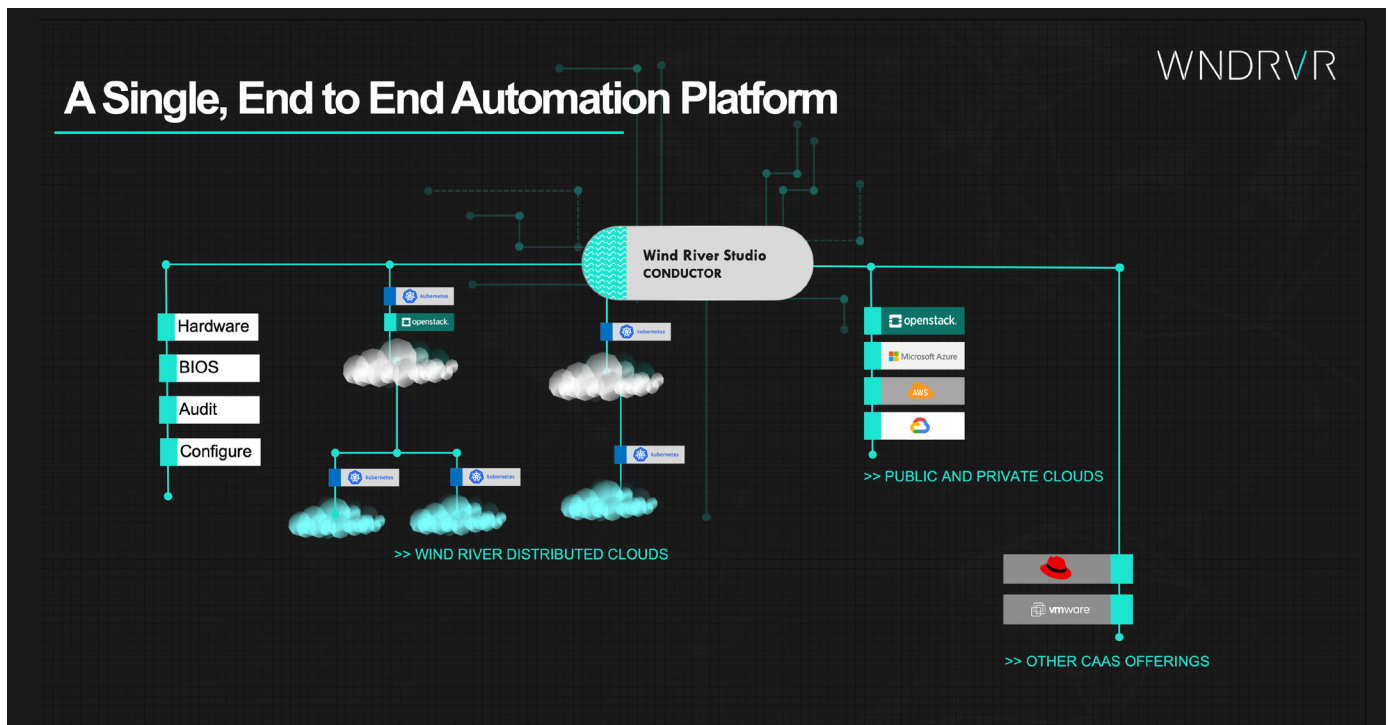### Key Capabilities of Wind River Studio Conductor

In the main research brief, we discussed how CSPs could improve edge success by selecting the best platform to host their edge applications, including vRAN and open RAN workloads. We also explained that the platform decision consists, at minimum, of a choice between underlying cloud and container platforms and a potentially separate choice around orchestration and management of workloads on that platform. Wind River Studio ties those two decisions together by offering an integrated solution containing both elements. While Wind River Studio Conductor primarily acts as orchestrator and automation system for the underlying Wind River Studio Cloud Platform, it can also automate and orchestrate other cloud platforms.

As part of this research brief, the sponsor of the brief, Wind River, requested that AvidThink provide an independent review of their solution on the topic of the research brief. AvidThink conducted this review using the information provided by the sponsor, and AvidThink's market research on edge computing. AvidThink has not verified all features and capabilities of the product. Readers interested in the solution under review should ascertain the veracity of vendor claims themselves. AvidThink cannot be held liable for the unexpected operation, damages, or incorrect operation due to any inaccuracies listed here.

The Conductor solution encompasses the following capabilities:

- **Multi-cloud orchestration** — Conductor is able to orchestrate other cloud platforms through plugins. Popular third-party platforms with available plugins include OpenStack, Kubernetes, AWS, Azure, Google Cloud, and VMware. Conductor can be further extended by coding plugins for other cloud platforms.

- **Single unified view of distributed components** — Conductor can display managed components and objects of varied types across distributed locations. Managed components include hardware and the cloud platform running on the hardware, but also middleware and application components.

- **Fully-distributed system support** — Conductor supports workloads in highly distributed environments, including single-site clusters likely to occur in edge deployments.

- **Model-driven Management** — Conductor maintains a model of the system under management and enables the creation and operation of workflows that operate on the model. This aligns with our recommendation to adopt an infrastructure-as-code approach and allows sophisticated management that is not possible with non-model driven systems.

- **Zero-touch automation for application deployment** — With thousands of edge nodes, zero-touch provisioning and automation must be extended from underlying hardware bring-up to application deployments (telco CNFs). Conductor supports advanced automation that allows for rapid distributed deployment of edge applications.

- **Secure foundations** — Conductor provides strict role-based access controls (RBAC) to ensure that administrators with appropriate privileges perform sensitive tasks. The platform also provides secrets management for the application workloads, preventing leak of information through compromised remote sites.

- **Integration to IT Service Management (ITSM) and DevOps Systems** — Enables operators to integrate ITSM ticket discipline as part of DevOps pipelines, and supports agile development processes at CSPs.

WIND RIVER STUDIO CONDUCTOR – MULTI-CLOUD AND EDGE ORCHESTRATION

(Source: Wind River)

When used in conjunction with Wind River Studio Cloud Platform, Conductor enables:

- **Remote bare-metal provisioning** — This capability provides for rapid bring up of large numbers of remote hardware devices.

- **Six-nines robustness and live updates** — Carrier-grade reliability ensures high uptimes and the ability to update the underlying cloud platform without downtime, including workload migration.

- **Efficient server core use** — Wind River's ongoing collaboration with Intel helps drive efficient resource usage on x86-architecture servers. For edge workloads, CSPs are looking to extract every ounce of performance out of the hardware, even as they are constrained by strict power and cooling requirements.

- **Massively parallel workflows** — Supports efficient day 2 operations including rapid upgrades and audits across large numbers of remote platform.

- **Complex declarative workloads** — Enables workload models that span multiple clouds (public and hybrid) and integrate seamlessly with external services.

AvidThink's Assessment of Wind River Studio Conductor

In our edge engagements and conversations with CSPs, there are active discussions on whether to separate the orchestration and cloud platform decisions or to make them jointly. From our perspective, selecting an integrated stack can reduce the risk for rollouts of performance and uptime-sensitive distributed workloads like vRAN and 5G O-RAN. It reduces platform friction and speeds time-to-deployment as well as time-to-market.

Wind River Studio has already proven its value in RAN deployments: Verizon, Vodafone, and now KDDI have made the decision to use Wind River as the foundation for virtualized and disaggregated RAN deployments. In some ways, this is not surprising given Wind River's reputation for real-time performance, deep understanding of time-sensitive workloads, and efficient system resource utilization. Elements that are important for vRAN/O-RAN workloads, and that bolsters our observations in the main research brief around MNOs using vRAN/O-RAN as one of the first edge telco-internal applications. This helps lay a foundation ready to support other near real-time or real-time applications for telco or enterprise use.

In comparing Conductor's capabilities, in concert with the Cloud Platform and Analytics, to the edge-native attributes required by CSPs, we note that it scores favorably across all key aspects. From being model based and automation centric to supporting hierarchical management and robust provisioning while having secure foundations, Conductor has done an admirable job in growing into an edge-friendly and edge-native solution.

AvidThink believes that the key to Conductor's long-term success in the orchestration market involves three elements. First, continued improvements in Wind River Studio Cloud Platform; since the integration with Conductor yields unique advantages that other pure-play orchestration vendors cannot easily mimic. Second, ongoing enhancements to its integration with third-party cloud platforms, particularly those of the hyperscalers. Third, as part of Wind River Studio, Conductor will need to help CSPs open up the enterprise edge platform market by easily allowing non-telco workloads to be run on premises, or in the network edge.

Wrapping Up

For CSPs looking to make strategic decisions around their edge cloud platforms, Wind River provides a compelling option in the form of Wind River Studio, including Conductor, the Cloud Platform, and Analytics. CSPs already embarking on vRAN/O-RAN trials and deployments will need to host performance and time-sensitive workloads; for them, Wind River's solution reflects a strength and heritage in real-time systems and offers a full-stack edge platform that is worthy of consideration.