White Paper

5G

Cloud Native 5G Core

Intel® technology and differentiators enabling the cloud native 5G core

Table of Contents

Table of Figures

Introduction

The cloud native 5G core is enabled by the intersection of multiple technology transitions across mobile networks, software architecture, and service delivery. The multiplying effect from the integration of these technology transitions will drive innovation and the new business models envisioned in 5G networks. The combination of a 5G radio connectivity with dynamic cloud architecture promises to unlock new innovative business models. However, the cloud native 5G core presents distinct requirements and unique challenges that must be addressed in the transition to the cloud model to allow for the desired return on investments in 5G radio. This document provides an overview of unique communications service providers (CoSP) challenges, cloud deployment models, characteristics of the cloud native 5G core, business models, and corresponding Intel initiatives and technology. The reader will gain baseline knowledge of these topics to help navigate engagements to have a meaningful dialogue on specific aspects of the cloud native 5G core and better understand Intel's comprehensive efforts and technology.

Cloud Native Deployment Models

Cloud Native Architecture Description and Benefits

Cloud native is an approach of building and running applications that fully exploit the advantages of the cloud computing delivery model. Cloud native is a new way that applications are developed, deployed, and maintained; it is NOT where the application resides (could be placed at edge, core, or data center cloud). A cloud-like delivery model is appropriate for both public and private clouds. Cloud native refers to decomposing apps into microservices. Some microservices might contain the application logic or might contain routing logic required to run the application. When these are disassembled it provides agility and flexibility to scale each component independently from each other which also provides an extra level of security. A container orchestration system (such as Docker or Kubernetes) provides the foundation for cloud-native architectures and symbolizes a new form of application architecture compared to what virtualization traditionally offered. Hardware-based virtualization provides a guest operating system for each virtual machine (VM), that provides application isolation from the host. Alternatively, containers consume less system resources as they use a common host kernel and OS. VMs provide an extra layer of isolation and security and containers can be deployed in VMs, which is common in public cloud applications. VMs are typically easier for operations teams to manage than bare metal (no hypervisor) environments. Containers, however, provide a more agile abstraction with more efficient resource utilization.

BENEFIT OF CLOUD NATIVE APPLICATION ARCHITECTURE	
Fast	Quickly create, update, and uninstall as needed. Request and provision more efficiently
Manageable	Containers take the complexity out of bundling, distributing, and installing applications.
Repeatable	Automated continuous development and integration. Create consistency between development, test, production, and deployment.
Portable	"Run Anywhere" Easily move software / workloads between the cloud, data center and edge
Increased Server Utilization	Shared resources and optimal workload placement to maximize server resource utilization.
Hyper Scale	Scalability to efficiently meet the need for additional compute or infrastructure resources.

A cloud native software architecture begins with the dynamic scalability and agility of the cloud in mind. The software is fundamentally architected differently from inception. Cloud native microservices **decompose** functions into modules with well-defined interfaces and operations.

Secondly, the microservice architecture is **packaged into containers** with all the necessary runtime requirements while sharing access to the operating system and infrastructure resources. For example, Kubernetes, an open-source platform for deploying containerized workloads, can be used to schedule the necessary cloud-native resources to complete the executable image including the supporting system tools and ecosystem building blocks.

Another key tenant of cloud native architecture is the **life cycle management** of the software. The continuous integration and continuous delivery (CI/CD) process refers to the ability to have targeted domains for new software and service validation. The CI/CD process automates the build, test, deploy process to expedite the transition from development to operations before being widely deployed at scale.

A cloud native architecture leads to continuous everything – thus increasing feature velocity, innovation, and revenue. A cloud native microservice architecture with an appropriately designed CI/CD pipeline is fundamental to realizing potential cost savings while also enabling business innovation. Appropriate design for ROI must incorporate optimal use of cloud resources by aligning workload requirements with infrastructure capabilities.

Automation and optimal workload placement for a service requires a detailed understanding of the infrastructure capabilities available to drive application performance and efficient resource utilization. Continuous capture of telemetry data is necessary to feed back into an efficient automation framework. The continuous telemetry collection from infrastructure, cloud stack and applications provide the necessary input data sources for intelligent machine learning (ML) algorithms. These ML/AI-based tools use this data to initiate informed decisions to provide proactive service assurance, intelligent workload placement and optimal power utilization to drive green data center initiatives.

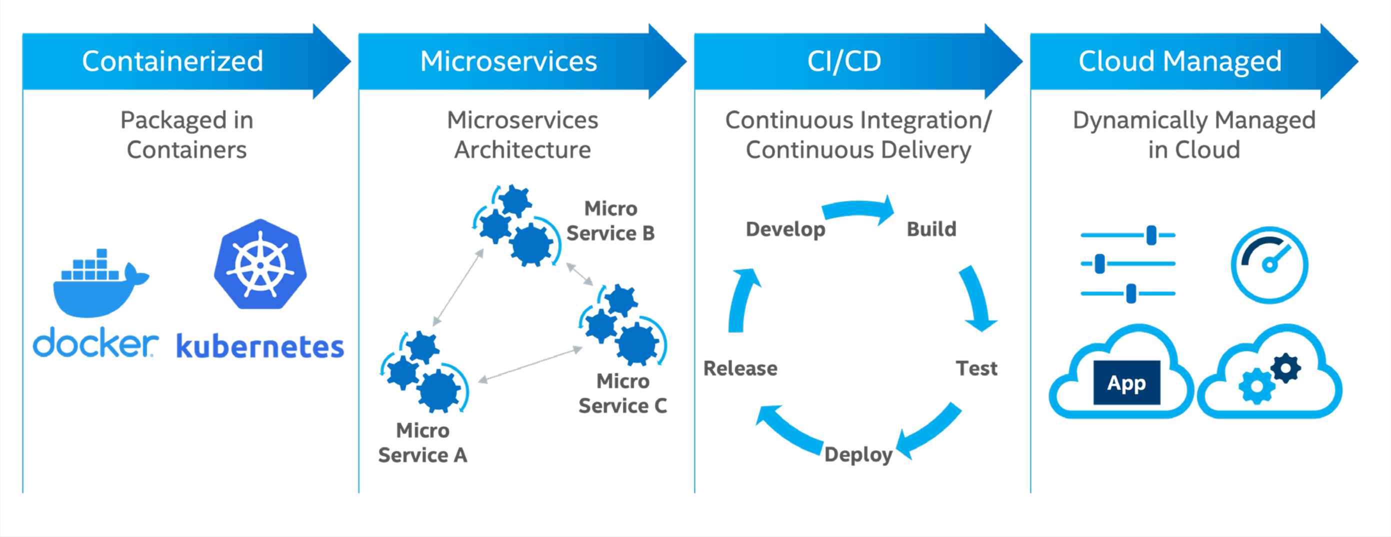


Figure 1. Cloud native journey.

Cloud Native Deployment Scenarios

Across the industry there is an abundance of cloud native deployment models and there is no single solution that works for all CoSPs that are planning to deploy cloud native 5G cores. This section delves into the considerations for different deployment architectures and methodologies across the platform as a service (PaaS), container a service (CaaS) and infrastructure layers. Given all the potential approaches and ecosystem partners, this paper identifies five basic deployment scenarios to allow for a focused discussion on the different models. The approaches are depicted in the diagram below.

- **Model 1:** Private cloud home grown by CoSP
- **Model 2:** Private cloud with virtualization/OS partners
- **Model 3:** Private cloud with telecom equipment partners
- **Model 4:** On-premises cloud with hyperscale partners
- **Model 5:** Private/public hyperscale cloud

For models 1 through 4, the CoSP provides the hardware with varying approaches and partners across the CaaS and PaaS layers. The 5th model is completely reliant on hyperscale partners. The containerized network functions (CNF) may or may not be developed by the CoSP. Traditionally, the CoSP is not the developer of the function. Deployment scenarios seeking to leverage CI/CD need to first identify how they plan to integrate externally developed CNFs into that pipeline and identify the lines of demarcation with the CNF suppliers.

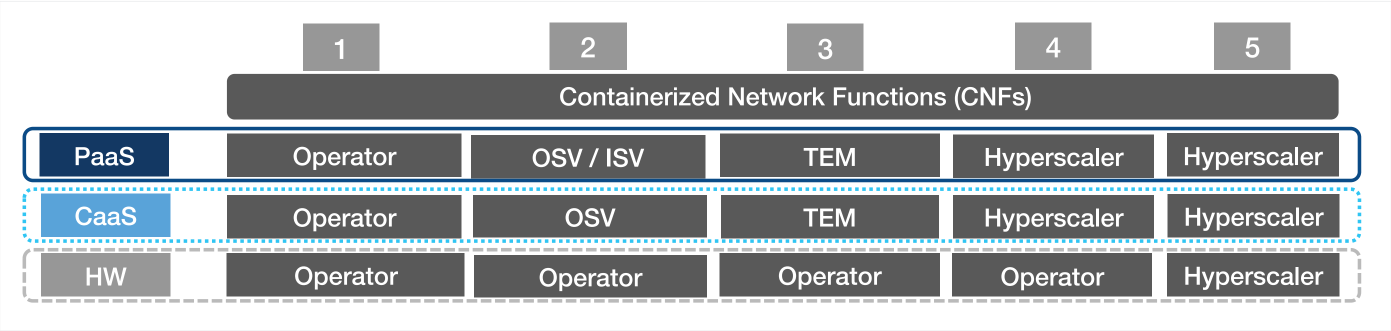
For both CaaS and PaaS, the platform deployment is not a user responsibility. The PaaS layer includes the software and hardware for an integrated application development and deployment environment. For PaaS, there is less operational overhead as the application is deployed automatically but the user gives up majority of control of how they are deployed. Google App Engine (GAE) is an example of PaaS.

The CaaS layer provides more control over scheduling and orchestration by means of a framework to deploy, run and manage containers and clusters using container-based virtualization. The CaaS manager monitors and handles failure detection and instantiation of new containers to provide service reliability. Google Kubernetes Engine (GKE) is an example of CaaS.

Regardless of the cloud stack approach and operational model, the same key tenants exist for an optimal cloud native 5G core must be considered:

- Performance and latency
- Security, service assurance and availability
- Power and operational efficiency
- Isolation and multi-tenancy

5G core service level agreements require careful evaluation and consideration of the deployment models and underlying operational controls that influence the characteristic identified. The deployment models and underlying resource utilization will correlate to the business value promised by cloud models.



	1	2	3	4	5
	Containerized Network Functions (CNFs)				
PaaS	Operator	OSV / ISV	TEM	Hyperscaler	Hyperscaler
CaaS	Operator	OSV	TEM	Hyperscaler	Hyperscaler
HW	Operator	Operator	Operator	Operator	Hyperscaler

Figure 2. Cloud native deployment scenarios

Cloud Native Considerations

The cloud native deployment model introduces unique system architectural considerations. A well-designed cloud native system architecture should be self-healing, secure, cost efficient, and leverage automation to be easily updated and maintained through CI/CD. CoSPs seeking to utilize CI/CD with externally developed applications need to consider how they implement CI/CD with their vendors. If architects fail to adapt their approach to these different considerations and not deploy an efficient system architecture the result may be fragile, expensive, vulnerable, and hard-to-maintain.

The table below highlights key cloud native architectural considerations.

DECOMPOSITION	The key characteristics of cloud native – DevOps, microservices, containers, and continuous delivery – rely on application decomposition. Decomposing the application reduces overhead and allows for focus on the key aspects of that service to drive increased value.
AUTOMATION	Although the upfront investment is often higher, favoring an automated solution will almost always pay off in the medium term in terms of effort, but also in terms of the resilience and performance of the system. Automated processes can repair, scale and deploy a system far faster than manual intervention.
STATELESS VS. STATEFUL	'Storing' of state is a careful consideration while architecting a cloud native architecture. Designing components to be stateless, wherever possible, architecting systems to be intentional about when, how and where to store state is critical for managing scalability, repair, rolling back and load balancing.
SECURITY	With a containerized cloud native approach, microservices communicate with each other via "service meshes" at both the data plane and control plane levels. A robust service mesh system must not only facilitate secure communications but also observe and identify suspicious behavior. Additionally, the heterogenous nature of microservices, distributed locations, hosting cloud environments, etc. presents complexities for addressing infrastructure security and maintaining the privacy and confidentiality of both the services and data.
ABSTRACTION	An infrastructure abstraction layer enables more efficient automation. Infrastructure that is hidden behind useful abstractions, controlled by APIs, managed by software, used by applications is necessary for managing that infrastructure in a scalable, efficient way.
POWER MANAGEMENT	Delivering performance in hardware and software to optimize workloads while maximizing power efficiency. Infrastructure telemetry allows for intelligent solutions to control and optimize power utilization across a data center.
HORIZONTAL BUILDING BLOCKS	The selection of hardware and software platforms that allow for ease of scalability and portability between clouds. Maximum scalability, flexibility and resiliency are key architecture considerations in the development of cloud-native applications.
MULTITENANCY	Multitenancy provides a layer of abstraction to drive value by sharing infrastructure resources with tradeoffs for performance and security challenges.
TOOLS	Testing, integration and deployment of packages require appropriate tools for automation that ensures CI/CD, scalability of the system, monitoring and recovery.
ELASTICITY	Using the right-sized amount of cloud resources and scaling in accordance with usage.
RESILIENCY	Maintaining the service level agreements for the service or customer.
COST EFFECTIVENESS	The cost effectiveness is dependent on the deployment model, the automation of the elasticity and resource efficiency (workload performance/cost).

5G Core Characteristics

The 5G core system architecture is specified as a **cloud-based system** that leverages a **service-based architecture (SBA)** to provide increased scalability and flexibility. Without an optimized 5G cloud native core, the investments in virtualized 5G RAN will not provide the desired performance benefits, operational savings, or enable the innovative services envisioned by the CoSP. **A cloud native 5G core must adhere to customer service level agreements for performance in addition to regulatory, privacy, monitoring, security, and auditing requirements within a constrained power and space envelope that provides a necessary return on investment.** This section highlights key characteristics to consider for an optimal cloud native 5G core.

The table below summarizes key characteristics of the 5G core cloud native functions.

5G NETWORK FUNCTION	USER PLANE PERFORMANCE	CONTROL PLANE PERFORMANCE	DB/DATA STORAGE	ENCRYPTION
User Plane Function (UPF)	x			x
Access and Mobility Management Function (AMF) Session Management Function (SMF)		x		x
Unified Data Repository (UDR) Unstructured Data Storage Function (UDSF)			x	x
Policy Control Function (PCF) Authentication Server Function (AUSF)		x		x
Service Communication Proxy (SCP) Security Edge Protection Proxy (SEPP) Network Exposure Function (NEF) Network Repository Function (NRF) Network Slice Selection Function (NSSF)		x		x
Security Gateway WebRTC Gateway	x		x	x

The 5G core system architecture consists of multiple network functions and introduces new functions to enable the service-based architecture and secure service exposure to external entities. The separation of control plane and user plane combined with a service-mesh architecture allows for true network programmability to enable ultimate deployment flexibility and independent scalability. Figure 3 highlights the key functions of the 5G system architecture.

The user plane functions (UPF) is the anchor point of the data session and demarcation point to the external data network. The packet processing **performance** and location of the UPF are key considerations for low latency and optimal application experience. The UPF is a complex, compute-intensive function that requires balancing compute offload, data plane acceleration, and efficient routing of packets to maintain the high network input and output. Multidimensional performance at scale across multiple vectors for packet inspection, rating, enforcement, encryption, enrichment, encapsulation, cache, and routing are required. UPF microservices must be architected and deployed to **leverage the infrastructure** capabilities for optimal packet processing performance and efficient resource utilization.

The 5G core system architecture allows for multiple UPFs for a given session while allowing for flexibility in UPF location to support low latency applications. The 5G Session Management Function (SMF) controls the UPF selection based on numerous criteria. This capability enables a distributed architecture and the ability to have "edge" deployments with the UPF residing near the application while the control plane remains centralized. When the UPF, or any 5G core function, is deployed outside the CoSP's trusted domain, there are unknowns related to both the infrastructure hardware and software stack. There needs to be assurances that the functions whether in private or public cloud are isolated and running in trusted execution environments.

Regardless of location, 5G core microservices must adhere to **regulatory, privacy and audit** requirements. Lawful interception (LI) is a regulatory requirement that allows appropriate authorities to perform interception of communication traffic for specific users. LI access must be considered when determining how to store and access data while maintaining an audit log. Cloud models are leveraging secure enclaves as a method to control access and manage keys to maintain data integrity with audit tracking for sensitive services and data across both trusted and untrusted domains. For the 5G system architecture, privacy and auditability apply to user data (UDM) and charging records that are generated by the 5G core functions. For example, the charging data records that are consumed at the charging gateway function must have a secure and auditable log of access to ensure the usage records are not manipulated.

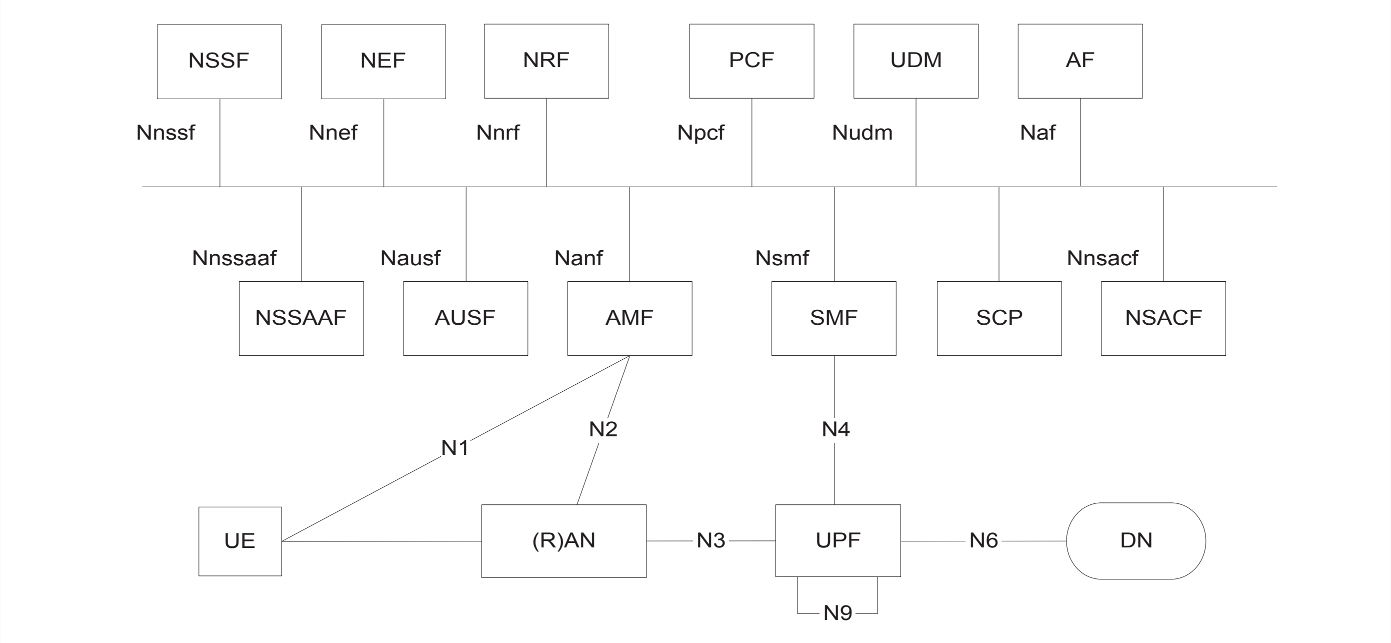


Figure 3. 5G system architecture (3GPP 23.501).

Security and **encryption** are prevalent throughout the 5G core service-based architecture. 3GPP requires encryption between the RAN and the core and confidentiality which can be addressed on the functions themselves or on an intermediary security gateway function. Private enterprise deployments also require encryption between small cell-based RAN. This encryption will increase with the rise of private 5G edge deployments. Leveraging infrastructure resources for encryption at scale is a key consideration during application development and during workload placement of the user plan functions. Careful consideration is also required for **monitoring and logging** across the distributed networks with unique encryption requirements.

The service-based architecture interfaces as shown in figure 4 provide a means for secure communication and authorization from different endpoints. The service mesh must be architected to provide efficient routing, load balancing, authentication, and monitoring. The use of sidecar proxies in a service mesh for each microservice consumes CPU and memory, which requires optimal routing and queue prioritizing to minimize latency impact. Key management for certificate authority signing and secure private key delivery are another security consideration for the 5G core service mesh.

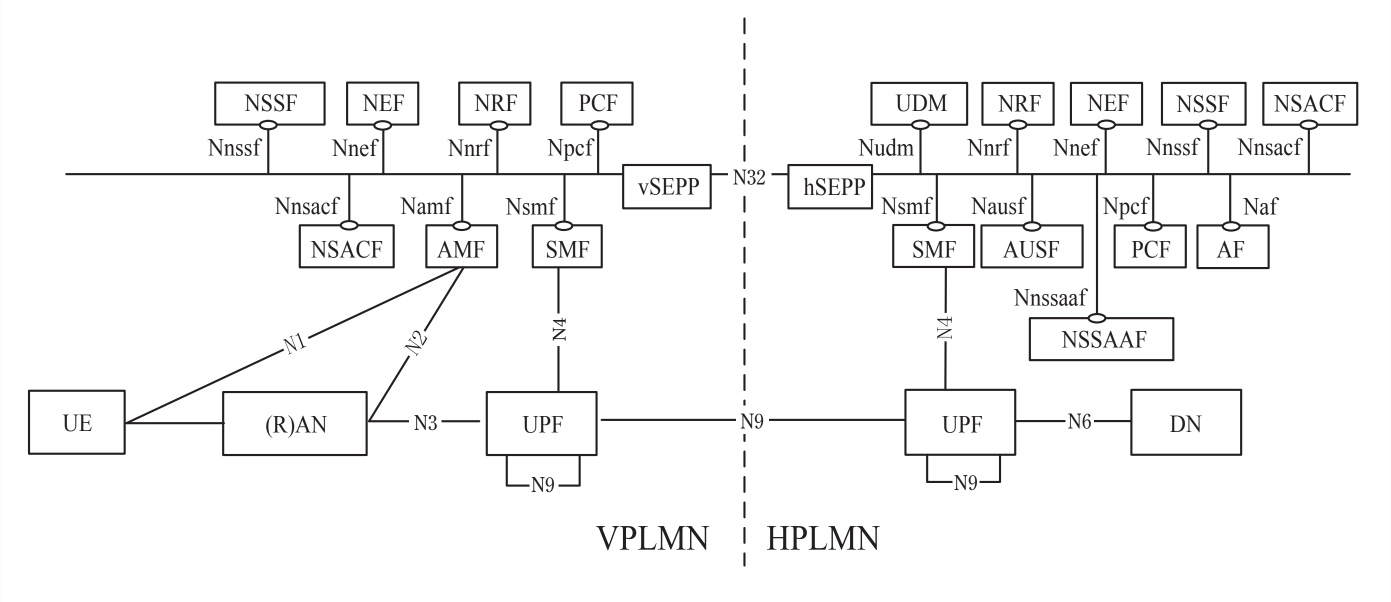


Figure 4. 5G roaming system architecture (3GPP 23.501).

The service-based interfaces (SBI) are also required to maintain **confidentiality and privacy**. While the security implementation and deployment approaches may vary for the SBIs within a trusted domain, it is an additional consideration factor when interfacing outside of the trusted domain for roaming scenarios. For example, the Security Edge Protection Proxy (SEPP) utilizes Transport Layer Security (TLS) to secure the interface for communications between the inter domain public land mobile network (PLMN) and the control plane. The network exposure function (NEF) uses TLS to provide confidentiality and privacy when exposing network function capabilities to external application functions.

An often-overlooked aspect of the cloud native 5G core is **storage**. 5G core network functions generate charging data records (CDRs) that must be securely stored and transported. As networks become more distributed, centralized storage of CDRs increases the cost of the solution related to the increase of transport of the data. Alternatively, edge storage is also a cost variable that will impact the economics of the solution. Optimized storage technology and architecture at the edge is a key component to reduce costs that impact the business model of any edge solution.

Another storage consideration is the how the uniform data repository (UDR) maintains and exposes the respective subscription profile, policy, and application data for the network functions services (UDM, PCF, HSS, etc.). Subscriber information must be available for the network to properly authenticate users, so **reliability** and security are paramount. How this data is accessed, transferred, and stored to ensure privacy with performance will require intelligent use of infrastructure resources across a distributed cloud network. As enterprises begin carving out network segments, techniques like database sharing that are used in today's multi cloud database distribution architectures may be utilized for scaling and coordinating private networks.

Business Modeling

The 3GPP 5G core standards define a service based architecture (SBA) that is designed from the start for cloud native 5G network deployment in which network functions communicate over open interfaces. Each network function provides a distinct set of services and exposes them to other network functions on the SBA. The result is a modular and highly adaptable core network that allows for the adoption of web scale technologies and software into telecom networks. This enables CoSPs to build and iterate network functions faster with greater scalability and flexibility.

Cloud native 5G paves the way for network slicing so providers can tailor network quality/performance to their client's specific service requirements. Network slicing promises to create new network-as-a-service (NaaS) opportunities and vertical enterprise solutions. CoSPs can optimize network performance and service requirements for enterprise users while maximizing their own limited spectrum and network resources. This will enable innovation and generate revenue streams through new B2B2X business models in which CoSPs provide wholesale services to business who then resell them to their own customers.

The new business motivation for moving to cloud native 5G core is to offer network APIs to innovative services that can leverage 5G capabilities such as network slicing, edge applications and fixed-mobile convergence. These offerings will create new business model opportunities to better monetize the network capabilities and generate service revenues while simultaneously optimizing the CapEx & OpEX.

While many of the innovations enabled by network APIs are yet to be defined, some of the more prevalent new business models in discussions include the following:

- **Neutral host access providers:** These providers leverage value in fiber access by divestment, re-configuring the communications value chain and providing fiber access to new CoSPs and edge entrants.
- **Mobile virtual network operators (MVNO) entering broadband market:** These providers can leverage neutral host fiber access in order to provide adaptive gateway functions (AGF) on edge cloud servers. The provider's goal is to deliver fixed and mobile services.
- **Wireline operator enters enterprise 5G market:** Wireline operators launching 5G mobile service offerings using fiber access networks must upgrade their CPE to be 5G capable in order to deploy MVNO (5G core) overlay and offer bundled fixed mobile broadband access to enterprises.
- **Cloud provider enters fixed and 5G mobile:** This operator lleverages neutral host fiber, deploys 5G CPE, and uses the BNG and AGF to enable fixed or mobile edge service breakout.

The innovation that will result from network-exposed APIs are wide and varied. The use of network APIs and web scale technologies will enable integrated solutions with modern technologies such as big data analytics, AI, and ML. These solutions will be used to optimize the network resources, understand customer behavior, and anticipate consumer needs to offer personalized and tailored services more efficiently. The integration of these technologies provides a multiplication effect on capabilities leading to powerful industry transformation driving new business innovations.

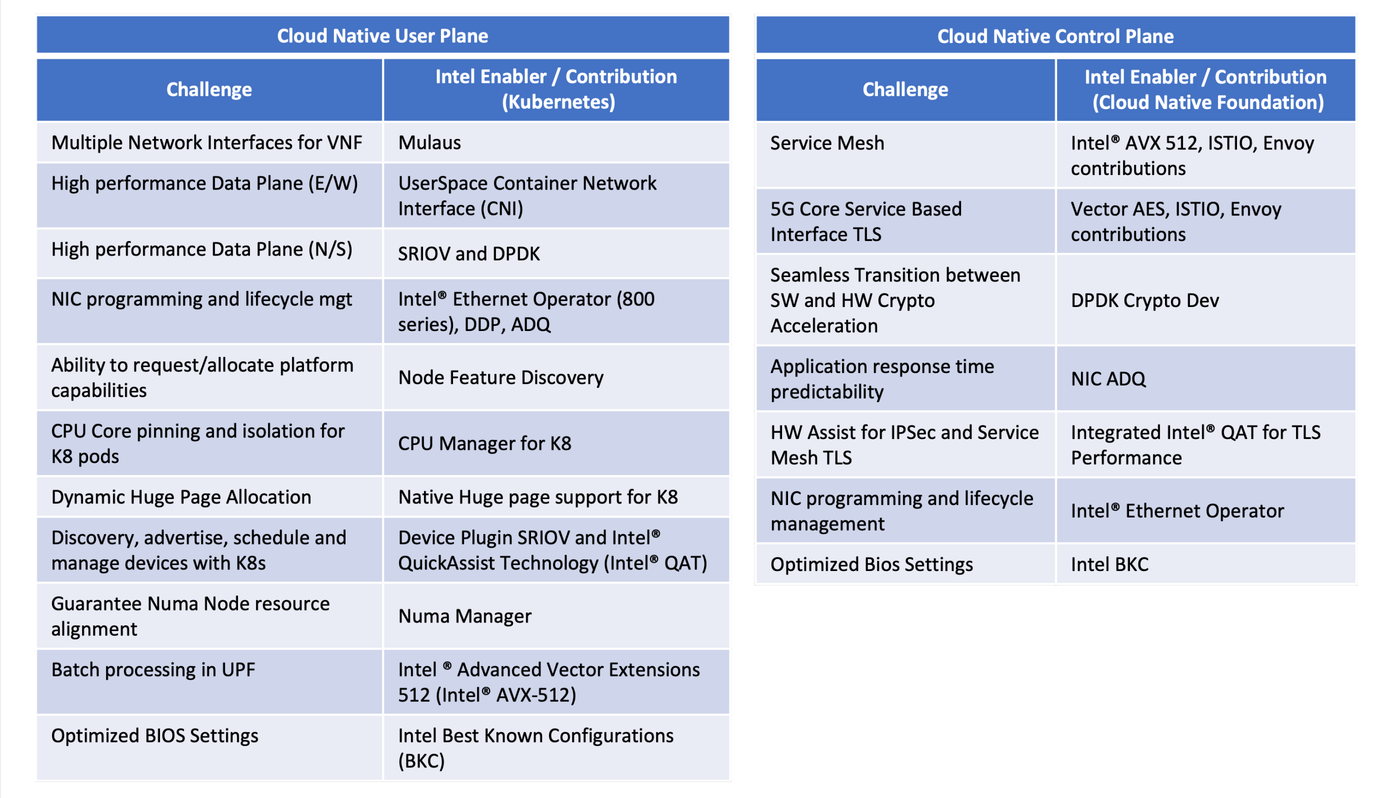
Intel Enabling Technology

Intel provides technology and ecosystem contributions to enable the cloud native 5G core. The tables below highlight specific technologies that are applicable to the unique characteristics of the 5G system architecture, and which are highlighted in previous sections of this document. For the sake of brevity, pointers to deeper dive collateral on each technology identified are available in the appendix of this document.

Intel enablers and contribution to Kubernetes for cloud native optimizations are highlighted below. For details and more information on Intel contributions to Kubernetes please refer to https://01.org/kubernetes.

5G CLOUD NATIVE CORE FUNCTION	INTEL TECHNOLOGY
Distributed Billing & Charging, Storage	• Intel® Software Guard Extensions (SGX) for secure and auditable CDR storage • Integrated Intel® QAT, AES-NI for encryption and crypto acceleration • AVX-512 for Crypto and database lookups acceleration • Optane byte-addressable low latency storage (distributed caching)
Service Based Architecture	• Integrated QAT for optimized TLS processing and crypto acceleration • AVX 512 for header compression and JSON data streaming • Intel® Software Guard Extensions (SGX) for key management in 5G Core Service Mesh • Istio contributions (several below)
5G User Plane (UPF)	• AVX2; AVX-512; AES-NI, other ISA for crypto acceleration • Vector AES (VAES) for IPSec; DPDK Cryptodev • Hyperscan for deep packet inspection • E800 Series Scheduler; E-DDP, Multus for optimized packet processing • SR-IOV; DPDK PMD; DPDK OVS; VPP for optimized packet processing • Kubernetes Contributions (see table below) • Optimized BIOS Settings • Linux kernel optimizations (isolcpus, huge pages, vCPU pinning, others) for optimal application performance and most efficient use of resources • NUMA optimizations (low latency UPI technology) • PCIe Gen4 NICs
5G Core Control Plane Optimization	• AVX2; AVX-512; AES-NI • Vector AES (VAES); DPDK Cryptodev • ADQ; NIC ADQ • Optimized BIOS Settings • Service Mesh and K8s
Sustainability	• Power Management: Cstates. PStates . granular controls for power consumption • Cloud Native Foundation, Telemetry for service assurance and workload placement
Ecosystem Contributions	• Ecosystem contributions (both open source and partner) • 3GPP, CNCF, other contributions and influence

Figure 5. Intel 5G Cloud Native Core Technology Enablers



Cloud Native User Plane	
Challenge	Intel Enabler / Contribution (Kubernetes)
Multiple Network Interfaces for VNF	Mulaus
High performance Data Plane (E/W)	UserSpace Container Network Interface (CNI)
High performance Data Plane (N/S)	SRIOV and DPDK
NIC programming and lifecycle mgt	Intel® Ethernet Operator (800 series), DDP, ADQ
Ability to request/allocate platform capabilities	Node Feature Discovery
CPU Core pinning and isolation for K8 pods	CPU Manager for K8
Dynamic Huge Page Allocation	Native Huge page support for K8
Discovery, advertise, schedule and manage devices with K8s	Device Plugin SRIOV and Intel® QuickAssist Technology (Intel® QAT)
Guarantee Numa Node resource alignment	Numa Manager
Batch processing in UPF	Intel ® Advanced Vector Extensions 512 (Intel® AVX-512)
Optimized BIOS Settings	Intel Best Known Configurations (BKC)

Cloud Native Control Plane	
Challenge	Intel Enabler / Contribution (Cloud Native Foundation)
Service Mesh	Intel® AVX 512, ISTIO, Envoy contributions
5G Core Service Based Interface TLS	Vector AES, ISTIO, Envoy contributions
Seamless Transition between SW and HW Crypto Acceleration	DPDK Crypto Dev
Application response time predictability	NIC ADQ
HW Assist for IPSec and Service Mesh TLS	Integrated Intel® QAT for TLS Performance
NIC programming and lifecycle management	Intel® Ethernet Operator
Optimized Bios Settings	Intel BKC

Figure 6. Intel 5G user and control plane Kubernetes contributions.

Ecosystem Partners and Solutions

An optimal 5G cloud native core requires an end-to-end service that is comprised of coordinated functions that provide performance, scale, flexibility, and operational efficiency. To reap the benefits of cloud computing and drive new business models with 5G services requires the foundational technical capabilities to be readily available in the ecosystem. Intel continues to work in coordination with OEMs, cloud providers, operating system partners, ISVs, and open-source projects to drive comprehensive solutions. Intel's efforts span the ecosystem driving innovation for cloud infrastructure, software stacks and operational tools that are necessary for an optimal cloud native 5G core. *These efforts are highlighted* in the diagram below.

The appendix has pointers to references demonstrating how Intel is enabling the cloud native 5G core transformation.

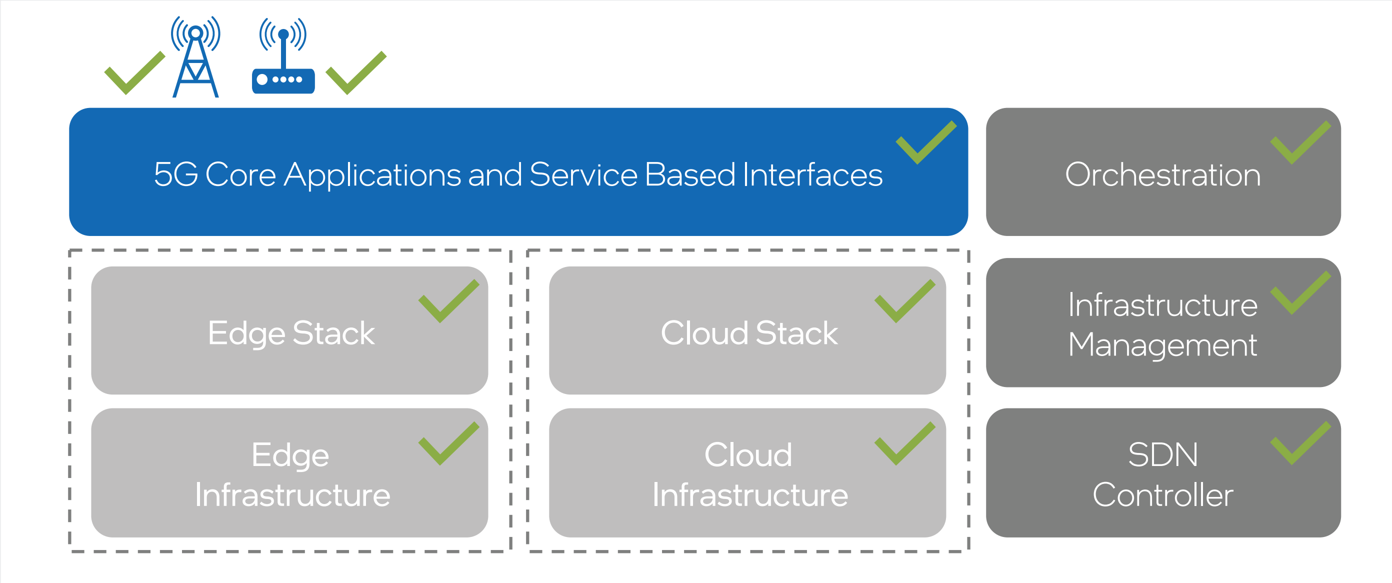


Figure 7. Intel ecosystem enablement.

Conclusion

The cloud native 5G core is a dramatic shift from previous mobile core architectures. The 5G system architecture fully embraces cloud methodologies and enables integration of modern software technology to revolutionize mobile core network function deployment models and operations. This dramatic change and decomposition of the system architecture will reshape the ecosystem landscape, driving new business partnerships and innovation. The characteristics of the 5G network functions and service-based architecture provide unique challenges for CoSPs that must be taken into consideration when embracing the cloud model. Intel is enabling the ecosystem with technology to deploy and operate cloud native, distributed 5G core networks to meet CoSP service level agreements while driving 5G solution innovation.

Appendix

Ecosystem Partner References

Below are some references demonstrating how Intel is enabling the cloud native 5G core transformation.

Google & Intel to accelerate cloud-native 5G

Intel & Metaswitch: Lighting Up the 5G Core

ZTE's High Performance 5G Core Network UPF Implementation Based on Intel

The Simpler Path to Cloud-Native 5G Networks: HPE 5G Core Stack, with Red Hat and Intel

Affirmed & Intel Enable Cost-Effective and High Scale 5G Core

Mavenir & Intel to develop products introducing a cloud-native paradigm to the network CoSPs

Rakuten Mobile, NEC and Intel demonstrate industry-leading performance in containerized 5G Core lab trial

VMware sees opportunity as CoSPs move from virtualization to cloud-native 5G

Red Hat and Intel blaze a more flexible path to 5G services, from hybrid cloud networking to edge computing

intel®

Notices & Disclaimers

Intel technologies may require enabled hardware, software or service activation.
No product or component can be absolutely secure.
Your costs and results may vary.

0222TM/H09/PDF ♲ Please Recycle 350447-001US