ORACLE
Communications

Converged Charging for Hyperscale Service Providers

Benefits of the Oracle Converged Charging System for the 5G Communications Era

Technical Brief

March 2021 | Version 1.00

SAFE HARBOR STATEMENT

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, timing, and pricing of any features or functionality described for Oracle's products may change and remains at the sole discretion of Oracle Corporation.

This document is provided for information purposes and should not be relied upon in making a purchasing decision. The contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose.

THIS DOCUMENT IS NOT PART OF A LICENSE AGREEMENT NOR CAN IT BE INCORPORATED INTO ANY CONTRACTUAL AGREEMENT WITH ORACLE CORPORATION OR ITS SUBSIDIARIES OR AFFILIATES.

Failure to adhere to these benchmarks does not constitute a breach of Oracle s obligations. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document.

TABLE OF CONTENTS

Safe Harbor Statement 1

Executive summary 3

Introduction 4

Ready for the Network, Ready for Business 4

The network charging endgame 4

Oracle's Converged Charging System for Communications 7

Mesh-Based Architecture, Webscale Charging Experience 7

High throughput, low latency 8

Near linear elastic scalability 8

Fault tolerant state and data management 10

Asynchronous Persistence and Revenue Management Integration 10

Cloud native deployment 11

Oracle Revenue Management Suite Cloud Native Deployment 11

Cloud Native Charging Grid 12

Scalability and performance test results 14

System Under Test 14

Observed Results 15

Aligned with the industry 16

Core Network Standards 16

Integrated Policy and Charging 18

TM Forum Open APIs 18

Close to the edge: a foundation for charging in the 5G era 19

Summary 21

Glossary of terms 23

EXECUTIVE SUMMARY

The 5G era presents new challenges for digital service providers to effectively monetize high volumes of communications, data and media traffic and at the same time provide a compelling customer experience. Modern charging systems will be required to deliver webscale operational efficiencies and at the same time function as a real-time experience engine for end users. The Oracle Converged Charging System (CCS), designed from the ground up to support the future needs of hyperscale service providers, is a cloud native 5G ready monetization platform that meets these needs.

Designed to operate at the intersection of core network and IT domains, the Oracle CCS uses mesh-based in-memory technology to provide high performance, resilient and linearly scalable charging, with pre-integrations available to advanced revenue management capabilities.

Cloud native, with a containerized and multi-service architecture, the Oracle CCS can take full advantage of innovations in cloud compute, storage, and networking technology to deliver an exceptional operational experience that can efficiently scale as your business grows. In a recent performance test conducted on Oracle Cloud Infrastructure with 50 million provisioned users, the Oracle CCS demonstrated single digit core charging latencies, high transaction throughput, efficient resource utilization and linear scalability.

Aligned with industry standards, spanning 3GPP and TM Forum Open APIs, the Oracle CCS is ready to help hyperscale communications service providers monetize the 5G future.

INTRODUCTION

Ready for the Network, Ready for Business

As 5G driven data volumes increase and customer expectations are raised, there is a compelling need to handle and process higher volumes of transactional charging data in real time.

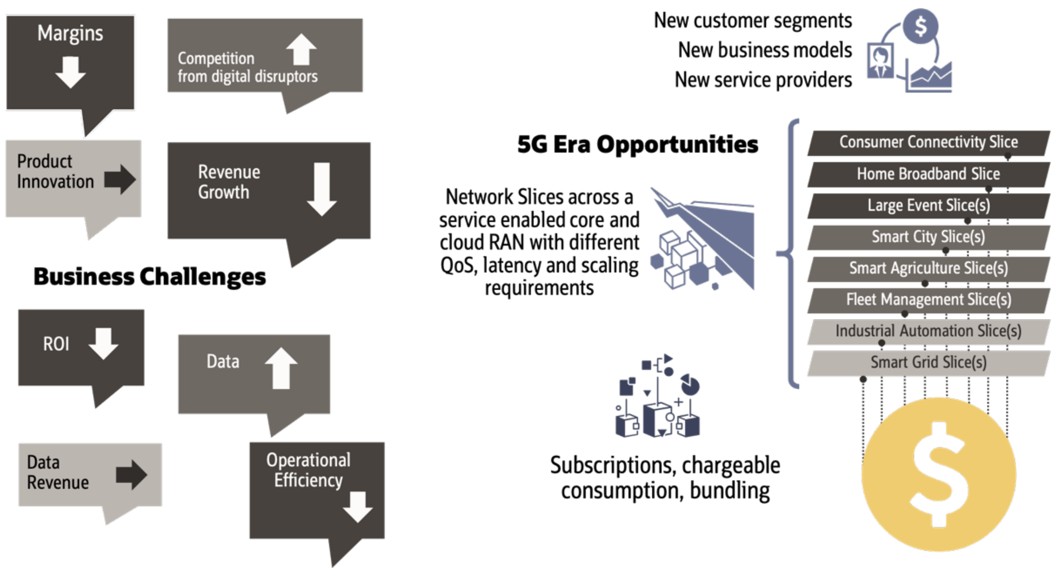


Figure 1 – Business challenges and opportunities in the 5G era

Some key business challenges include:

- Ensuring accurate data charging and transactional consistency, with the charging system introducing no revenue leakage
- Overcoming traditionally fragmented business systems, processes, and channels
- Providing a consistent view of customer types (consumer, enterprise, prepaid, and postpaid)
- Driving a compelling user experience
- Supporting future 5G charging model complexity
- Supporting low latency for service authorization and delivery
- Delivering cost efficient, high performance charging (extreme volumes, complex account structures, and many devices and sensors)
- Supporting high availability ("always on") deployment
- Taking optimal advantage of modern cloud compute infrastructure (aligned with new and future deployment architectures)
- Aligning business systems with new organizational structures that span the IT and network domains

The network charging endgame

The evolution of network charging for mobile communications services has had some significant events over the past few decades, driven by the changing services landscape, new network technologies, and the rise of commodity broadband technologies driving new user demands (figure 2). Prior to the rise of mobile prepaid in the early to mid 1990s, postpaid billing was the only way for CSPs to generate revenue, and they used offline charging (post processing of network CDRs) to generate invoices.

The rise of mobile prepaid services was initially conceived as a revenue assurance mechanism to mitigate fraud and bad debt, barring the service if the user did not have enough credit. Early "hot billing" prepaid systems were essentially modified postpaid billing systems, which sent a barring command to the Home Location Register

(HLR). However, the fundamental issue with this approach was that nothing was real-time; there was a delay between the CDR being generated and the activation of barring in the HLR during which calls could essentially be made for free (both intentionally and unintentionally).

This industry challenge led to the development of the service node platform, where there was a real-time interaction between the mobile core network and the charging system. This approach removed the window of negative credit and was the origin of the online charging system. As a result, the prepaid service was no longer a system for customers unable to pay, but an amazing opportunity for the service provider to launch highly marketable offerings where the end user was in complete control of their spend. The prepaid boom began and has not slowed down since, representing the largest segment of mobile consumer subscribers for the past two decades and continuing into today's 5G era.

As prepaid systems evolved through the use of IN protocols such as INAP (Intelligent Network Application Part) and CAP (CAMEL Application Part), there was a divergence in the ownership of prepaid systems (the network domain) and the ownership of postpaid billing systems (the IT domain). This organizational split has grown over the decades and is still prevalent in the industry, reflected by the different types of charging and billing platforms available in the market today.

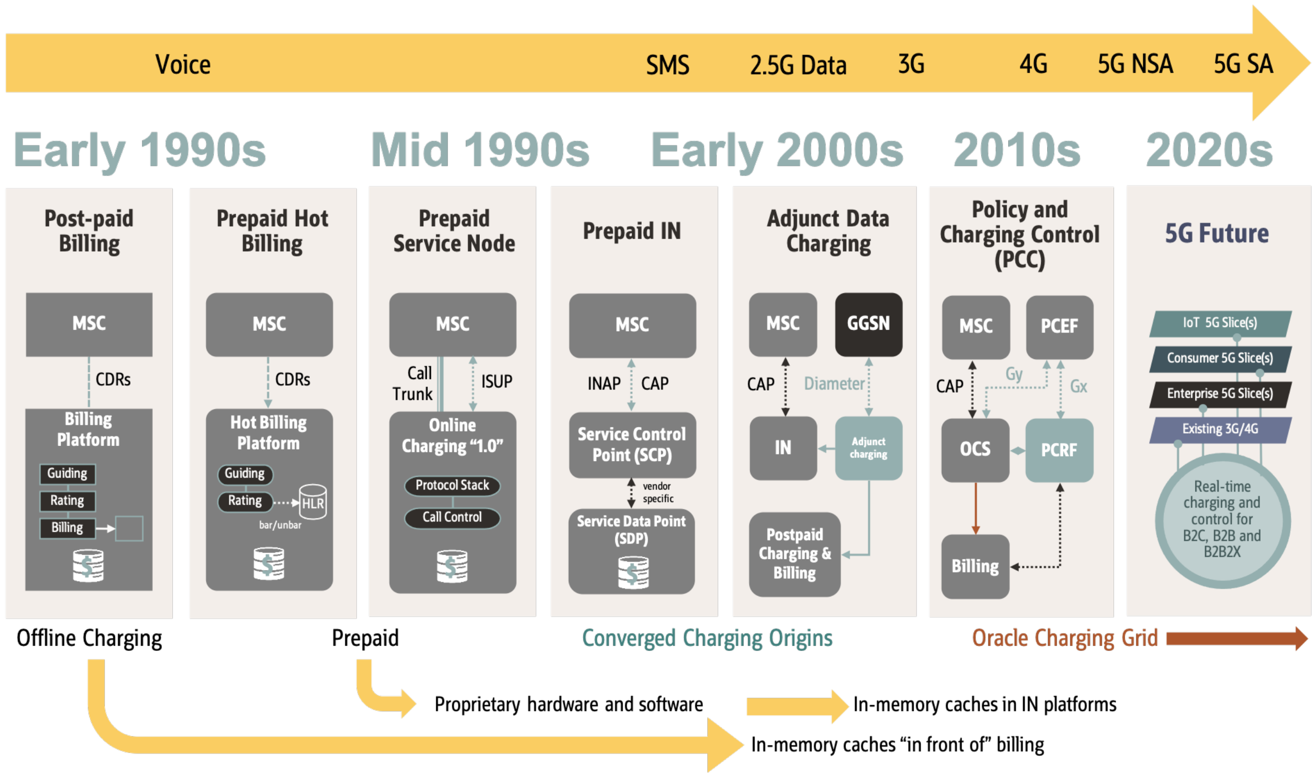


Figure 2 – The evolution of network charging

Around the late 1990s/early 2000s the concept of "convergent charging" (now referred to as "converged charging") started to be discussed in the operator community as the ideal end state vision for charging systems; applying the value of online charging to both prepaid and postpaid users. The benefits of real-time spend management and control could, in principle, now be applied to postpaid users, for example through the setting and triggering of spend thresholds resulting in real-time notifications that helped to mitigate "bill shock".

The term "convergent charging" has been interpreted on a number of levels: the convergence of prepaid and postpaid in a single charging architecture, support for blended multi-service offers with combinations of prepaid and postpaid charging options, and at the higher-level support for multiple network types and converging industry domains. However, many CSPs were presented with a resistance to change based on their existing organizational structures and the significant charging and billing investments that had evolved over the years. As such, for some operators converged charging remained more vision than nearer term practicality.

The rise of mobile data services led to a dramatic increase in the volume of events and also presented challenges to traditional prepaid focused IN platforms with regards to supporting non-linear, volume-based data charging. IN platforms evolved over time to meet these new demands, incorporating newer in-memory database cache technologies to support lower latencies and higher performance. Similarly, postpaid systems started to incorporate in-memory architectures in a move towards realizing the converged charging vision of a true digital user experience with contextual notifications and real-time balance visibility.

These approaches of evolving IN platforms and postpaid billing systems dominated the charging industry in the 2000s, with some alternative approaches emerging that offered new types of “pure play” charging systems intended to be integrated with or work alongside existing systems. However, the rise of 4G and now 5G, with its IT cloud native service-based architecture, has caused service providers to rethink not only their architectural vision, but also their organizational structures.

The boundaries between network and IT services are now merging, and the approaches of incrementally enhancing existing IN and billing system architectures or paying for costly integration of standalone charging engines will no longer be enough to support the 5G digital demands. In order to truly deliver on the original goal of converged charging, 5G era monetization systems will need to combine network grade performance and scalability with IT cloud native capabilities needed for modern operations, alongside pre-integrations to advanced billing and revenue management capabilities (figure 3).

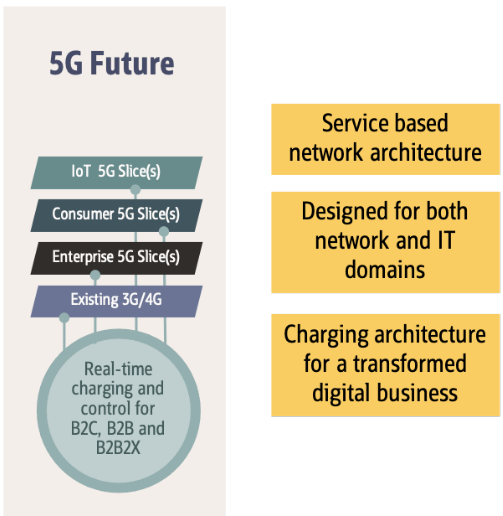


Figure 3 – 5G converged charging requirements

ORACLE'S CONVERGED CHARGING SYSTEM FOR COMMUNICATIONS

Oracle's Converged Charging System (CCS), powered by industry leading in-memory grid technology, has been designed from the outset to support the technical and business monetization demands for hyperscale 5G digital communications providers. It is a real-time transactional system of record for converged data and communications session charging and balance management, 3GPP aligned with native integration into Oracle's full suite of billing and revenue management capabilities designed in accordance with TM Forum principles.

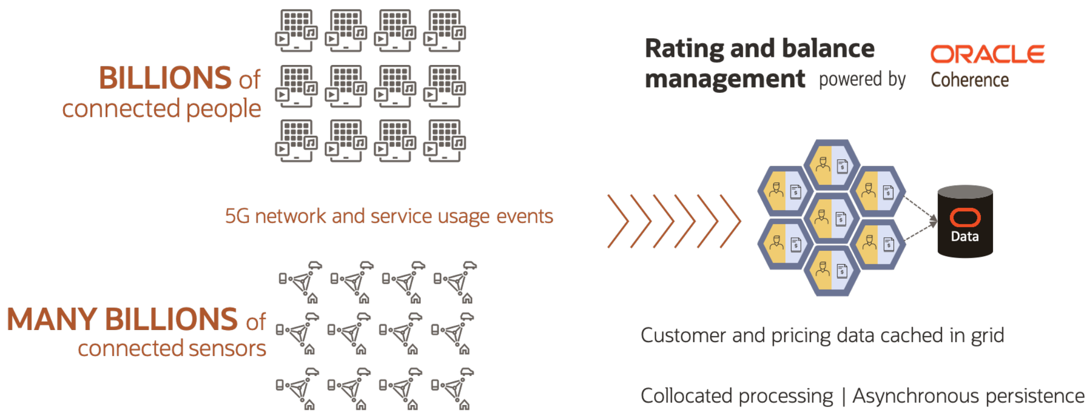


Figure 4 – Oracle's charging grid Technology

Built around network and IT industry standards, the Oracle CCS uses an innovative high performance and coherent data management architecture to support near linear scalability, low latency, and highly available multi-site deployment with transactional consistency (figure 4).

Mesh-Based Architecture, Webscale Charging Experience

The real-time rating and balance management functions are underpinned by the industry leading Oracle Coherence in-memory data grid technology, forming a high performance and resilient **charging grid** which enables a webscale experience. Coherence has a **dynamic mesh-based architecture** that provides fast data access and enables predictable scalability for mission critical applications.

The use of in-memory technology in modern network charging applications is essential to support the very low latency service authorization and re-authorization network requests required, typically specified in the order of milliseconds.

However, simply supporting low millisecond latency in itself is not enough to satisfy the key non-functional scope of a 5G ready network charging and control solution. Tomorrow's charging systems must support high throughput parallel transaction processing, near linear scalability, and highly available geo-redundant deployment models, alongside always accurate rating, balance and threshold management for complex product offerings supporting multi-device ownership structures and account sharing models.

The Oracle CCS charging grid adopts an innovative approach that co-locates the processing and data, offering high degrees of parallelism, with events persisted asynchronously to an enterprise class database ensuring efficient processing and low latencies (figure 5).

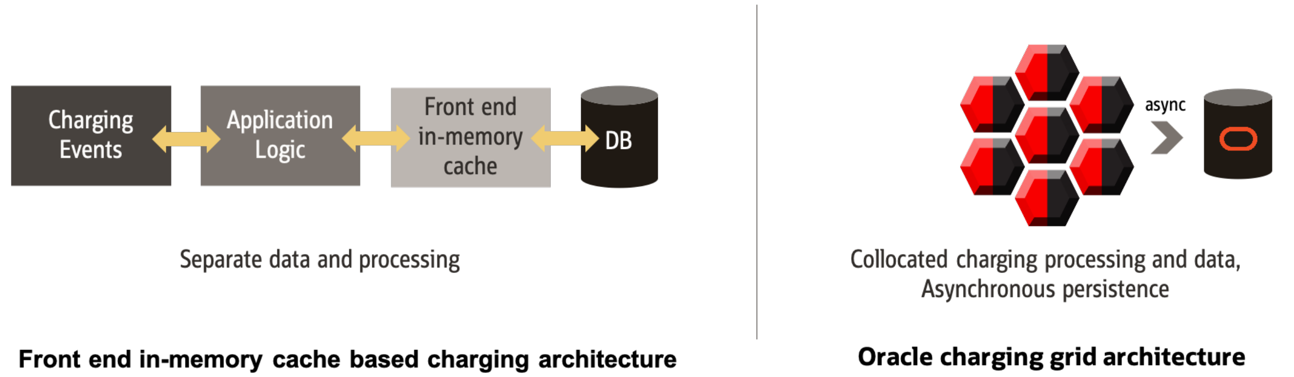


Figure 5 – Charging system in-memory architecture comparison

High throughput, low latency

The Oracle CCS charging grid stores customer and pricing data using in-memory cache technology distributed across a cluster of grid members (realized as JVM nodes), with data entries serialized in key-value pairs. Read and write latencies are extremely small, supporting very low end-to-end charging transaction response times for data session initiate, update, and terminate requests. The Oracle CCS uses Coherence distributed caching for storing customer objects across members of the charging grid with automatic partitioning and rebalancing of data as new members are added or removed from the grid.

Rather than taking the approach of fetching data from a remote store, performing processing, and then writing the data back to the remote store, the Oracle charging grid processes all charging transaction requests directly where the data entries are managed in the cluster. This co-located data and processing affinity architecture offers the following benefits:

- Processing is extremely fast as all objects are held in-memory, ensuring low latency and cost-efficient compute resource utilization
- Data access times are close to zero, with processing invoking optimized HashMap lookups
- Almost zero cost locking, retaining transactional data consistency and ensuring no revenue leakage within the charging system

The pricing domain model is based on the industry standard TM Forum Shared Information and Data Model (SID).

All customer data objects are stored in a distributed (or partitioned) fault tolerant cache, which is spread across all the grid servers in such a way that no two servers are responsible for the same piece of data. The distributed cache model supports linear scalability and high performance.

All designed rate plans are implemented in memory as optimized run-time rating graphs and supported by the Coherence Continuous Query Cache model.

Near linear elastic scalability

The grid is fully distributed, with no single point of contention, supporting independent scalability for large and growing customer data sets. The Oracle CCS charging grid supports near linear scalability due to the automatic partitioning of customer data objects across the grid members (figure 6). Coherence detects new grid members and automatically re-balances the cache data so that it is spread evenly across the grid. In distributed caching, read operations are only ever a single hop at most from the charging data, ensuring a scalable model that can take optimal advantage of switched network infrastructure.

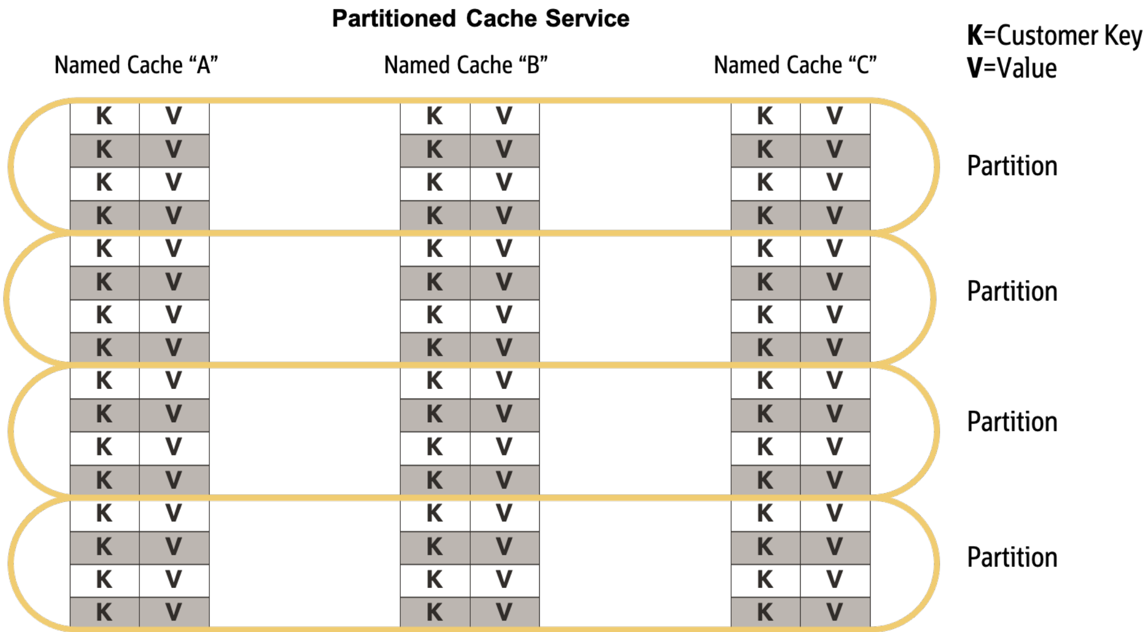
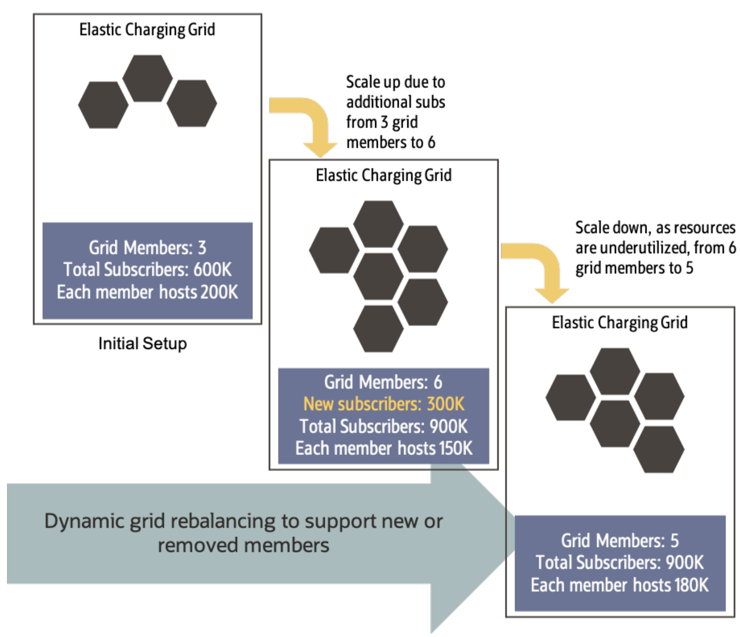


Figure 6 – Automatic partitioning of customer data

Dynamic scaling up or down can be handled "in-flight" to support changes in presented traffic load, subscriber growth, or compute availability, for example to change roles between test and production compute for efficient resource utilization (figure 7).



- Coherence based charging grid
- **Collocated data and operations** for increased efficiency and throughput, with low latency
- Configured to **keep-a-copy** for high availability of data
- Supports **scaling up or down** dynamically
- Grid **rebalances** automatically every time a new member is added or removed to distribute subscribers evenly across the grid members

Figure 7 – CCS elastic scaling model

Fault tolerant state and data management

The Oracle CCS charging grid supports fault tolerant state and data management. Using the Coherence distributed cache mechanism, changes to customer data are automatically replicated in real-time to one or more backup nodes to ensure high availability with almost zero impact on node failure (figure 8).

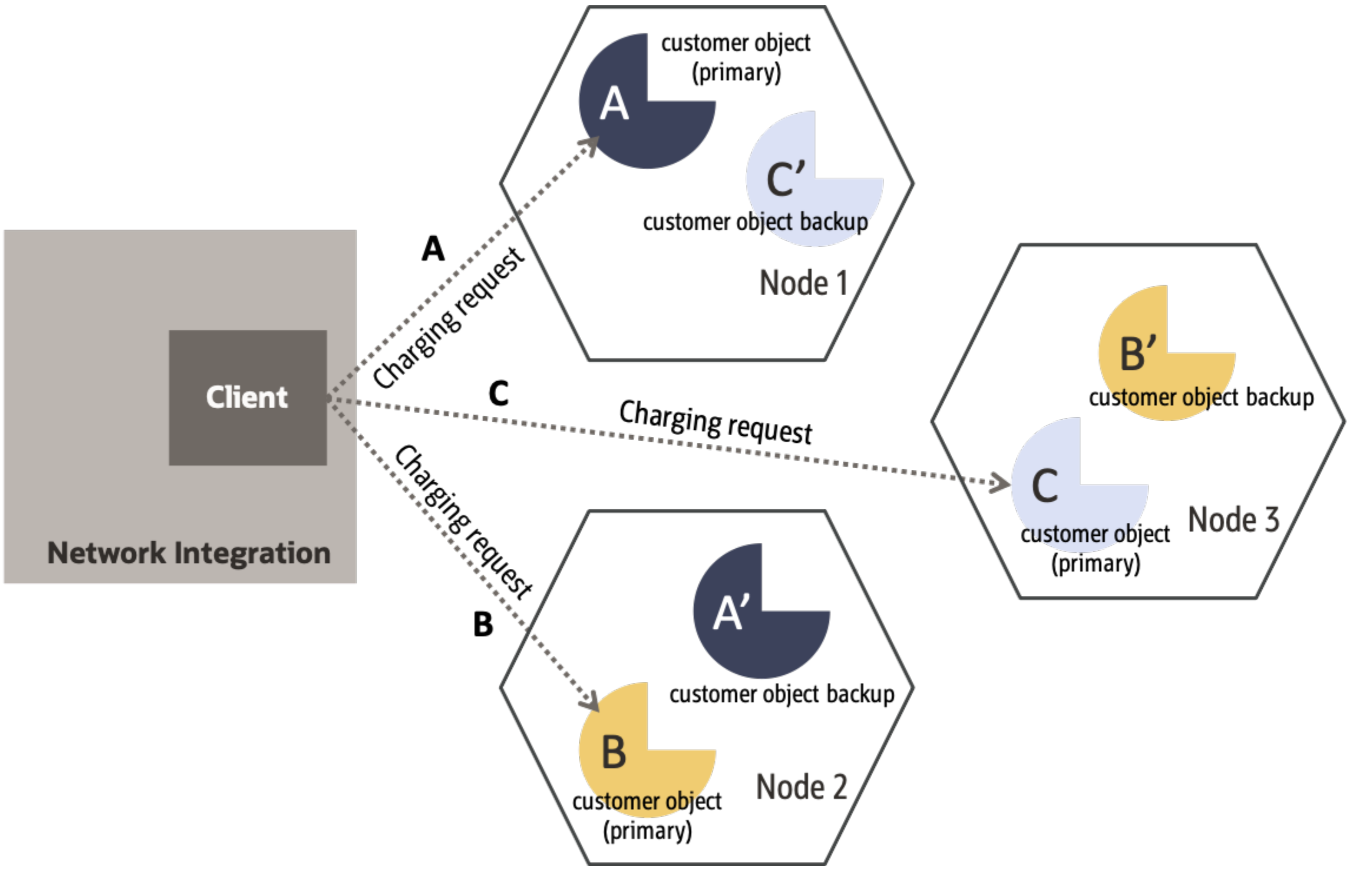


Figure 8 – Charging grid cluster backups

To support geo-redundant high availability scenarios, multiple deployment options are available to align with specific business objectives, the most comprehensive option being **multi-site active-active deployment**. This model takes advantage of Coherence Federated Caching, which permits a multi data-center cluster to replicate data over a WAN interconnect.

Asynchronous Persistence and Revenue Management Integration

Asynchronous persistence of the grid cache ensures high performance without compromising business-critical data availability. Rated events are offloaded asynchronously to revenue management functions providing a near real-time event flow that does not impact the core network charging processing. Off grid persistence of customer account and pricing data is stored in an Oracle database that underpins a complete suite of pre-integrated billing and revenue management applications.

CLOUD NATIVE DEPLOYMENT

Modern monetization systems will need to take optimal advantage of compute, network, and storage infrastructure to operate and scale efficiently and grow as the business demands. These requirements translate to the need for monetization systems to support a cloud native containerized, orchestrated, and multi-service deployment architecture.

5G is enabled by a new Radio Access Network (RAN) called the 5G New Radio (NR) and a new core network called the 5G Core (5GC), linked by high-speed backhaul connectivity. It is a revolutionary system architecture that applies webscale IT principles to a telecommunications core network. The sheer scale, performance, and efficiency required to support large scale, low latency data sessions across millions of personal devices and potentially trillions of sensors have driven the need for a core architecture that can take optimal advantage of modern innovations in compute, network, and storage technologies to support efficient scaling, high-transaction throughput, security, and resiliency.

The 5G SA model has a new Service Based Architecture, standardized by the 3GPP Release 15 and 16 specifications and based on decoupled network functions that communicate with each other using RESTful service invocations between consumer and producer over the HTTP/2 protocol. This is a fundamental shift from more traditional network protocols such as Diameter (the core language of 4G EPC network signaling) which lends itself to **cloud native network function deployment.**

A key principle of the 5G Service Based Architecture is the separation of the user plane and the control plane. This separation enables core network functions to be efficiently scaled independently from the user plane. If the 5G core network has an IT-based, cloud native architecture, then the converged charging system should as well, ensuring operational consistency.

Oracle Revenue Management Suite Cloud Native Deployment

The Oracle Converged Charging System is a key component of a complete Revenue Management Suite, built around a multi-service architecture that can be deployed in a cloud native environment as a family of orchestrated containers (figure 9, depicting a functional view of the suite). The CCS can be configured to run as a cloud native application in a containerized and orchestrated deployment architecture, taking advantage of cloud native infrastructure and DevOps CI/CD tooling to enable service providers to design, test, and deploy services more quickly, operate more efficiently, and scale as business needs grow.

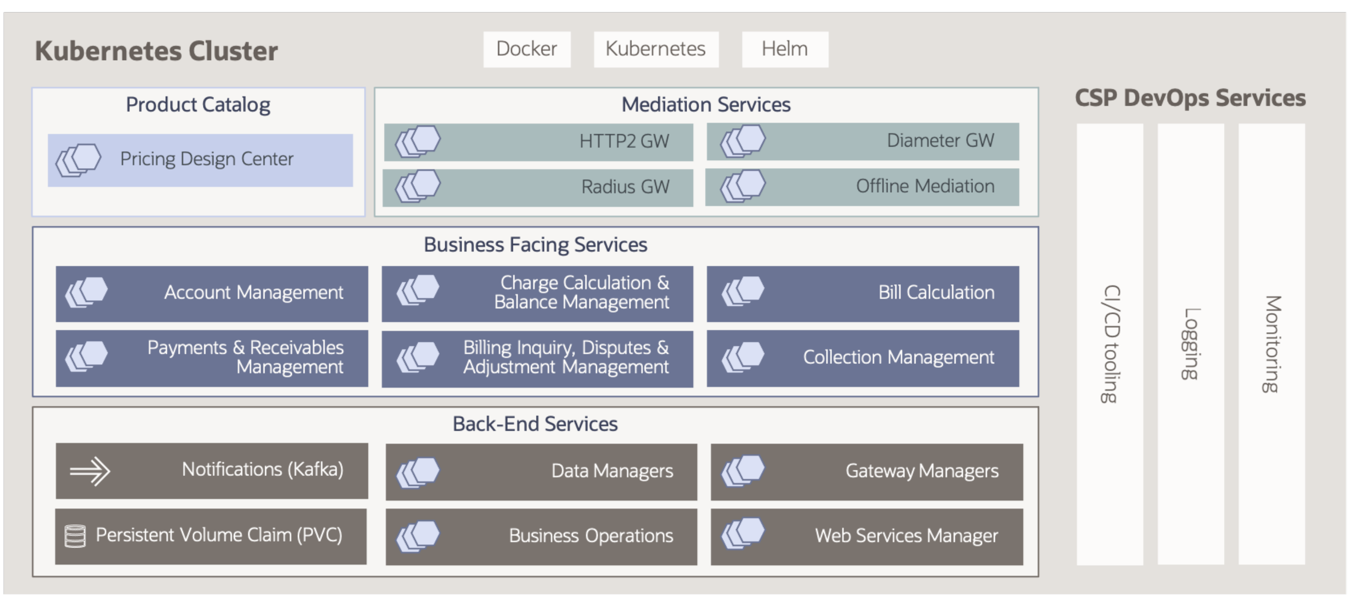


Figure 9 – The Oracle cloud native revenue management suite – functional view

The Revenue Management Suite has a multi-service architecture with each service provided as a Docker image for deploying as a run-time container in a Kubernetes cluster on cloud infrastructure. This deployment option takes advantage of industry accepted cloud native technologies including Kubernetes for container orchestration, Helm for packaging and deployment, and the EFK stack, composed of ElasticSearch, Fluentd, and Kibana for logging.

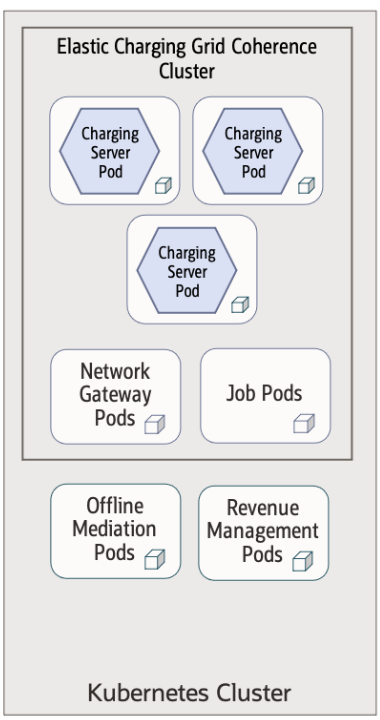
Cloud Native Charging Grid

The benefits of the charging grid previously discussed are significantly enhanced with the Oracle CCS cloud native deployment model. The enhanced benefits include:

- Optimized installation time
- Rapid environment replication for development and testing and faster root cause analysis of potential production issues
- Self-healing capabilities for greater service availability
- Simpler updates
- Efficient scaling that takes maximum advantage of the available compute resources (nodes)
- Faster launch of market offerings by taking advantage of CI/CD toolchain integration and automation

The charging grid architecture provides a reliable and scalable cloud native platform for charging state management and high-volume processing. The core charging server and network gateway pods are deployed as Kubernetes StatefulSets, which enables simple declarative scaling by increasing the replica count to support subscriber growth and increased network traffic loads. When the replica count increases, a new empty charging server pod is created which Coherence then automatically rebalances with the customer data objects stored in the distributed cache. The partitioned data is replicated across a configurable number of grid members for backup as previously described.

Figure 10 shows the key concepts of the cloud native charging grid architecture, with figure 11 illustrating Coherence Cache replication in a multi-site deployment.



- Kubernetes manages the grid members as Pods
- Charging Server and Network Gateway pods deployed as Kubernetes StatefulSets
- **Kubernetes managed scaling**
  - **Scale up** adds a new Pod, which in turn adds a new member to the grid
  - **Scale down** removes an existing Pod, which in turn removes the member from that Pod
  - Addition or removal of a grid member results in **rebalancing** of grid members
- Charging grid Pods coexist with other Pods in the Kubernetes cluster

Figure 10 – The Oracle cloud native charging grid

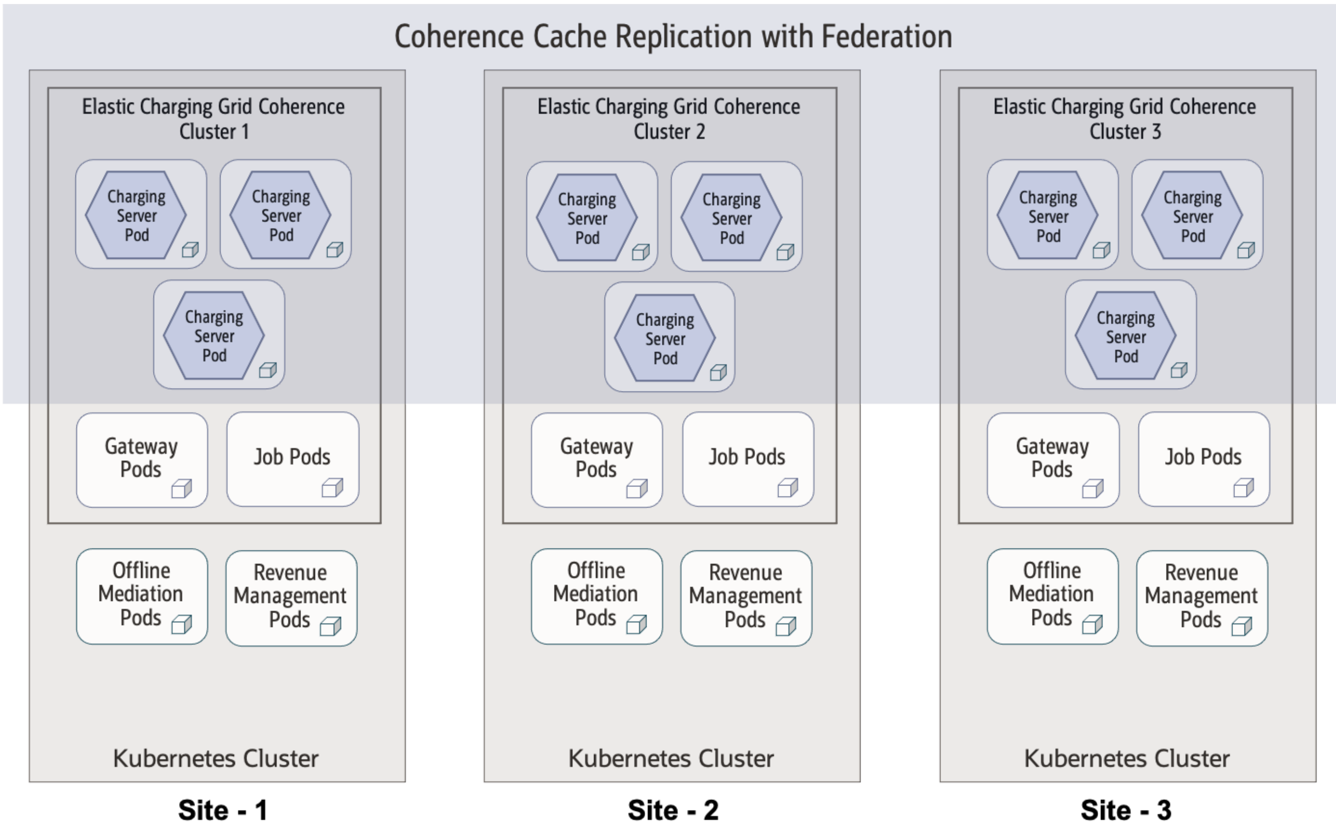


Figure 11 – Charging grid multi-site deployment model

The use of industry standard cloud native tooling provides improved efficiency for regular CCS operations to ensure smooth running of the system. For example, all system pods are able to be integrated with the ElasticSearch, Fluentd, and Kibana (EFK) stack for centralized logging and visualization.

Each Fluentd agent pod is deployed within a Kubernetes DaemonSet, which deploys one pod for each node in the application cluster. The CCS application pods write logs to `stdout`, which are collected by Fluentd which persists them in the ElasticSearch object store, ready for visualization by Kibana. The Oracle CCS can also be deployed with an alternative logging technology of the operator's choice if required.

Oracle's approach to developing cloud native business support systems builds on decades of experience developing industry leading monetization and orchestration products for digital service providers. The solution supports mission critical functions, capturing and delivering revenue for the business and as such should be extensible, flexible, and future ready, evolving ahead of the demands of the industry to provide a growth platform for all of our customers. The Oracle CCS cloud native charging grid architecture builds on Oracle's data management and cloud native innovations, providing an ideal platform for future network automation requirements as the industry moves increasingly towards highly distributed, low latency, agile environments with zero touch operations.

SCALABILITY AND PERFORMANCE TEST RESULTS

In a recent test conducted on Oracle Cloud Infrastructure with a provisioned test subscriber base of **50 million mobile subscribers**, the Oracle Converged Charging System demonstrated compelling performance and scalability characteristics.

System Under Test

Mixed charging traffic generated from a core network simulator was presented to a cloud native CCS deployment with 50 million provisioned accounts, consisting of 90% prepaid and 10% postpaid user profiles. The charging traffic simulated a mixture of initiate, update and terminate charging operations across data, voice and messaging, including balance enquires and notifications.

Cloud Infrastructure Hardware

The test environment, hosted on Oracle Cloud Infrastructure, is depicted in figure 12.

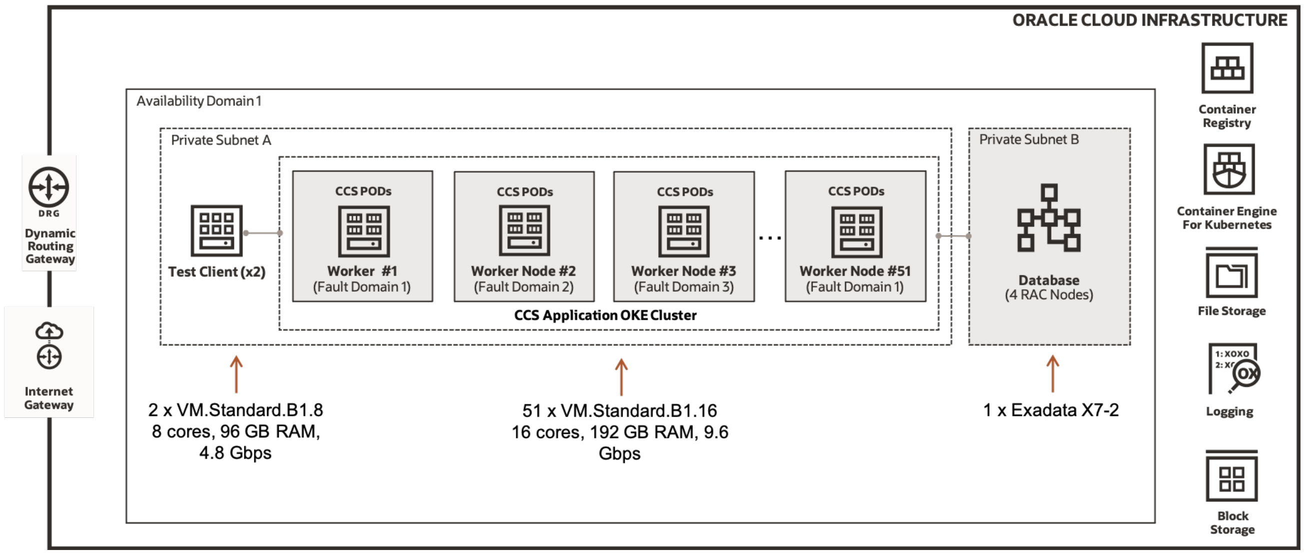


Figure 12 – Scalability and Performance cloud infrastructure test environment

The test driver, generating charging operation requests, was deployed on two standard 8-core virtual machine shapes (VM.Standard.B1.8), with the Oracle converged charging system Kubernetes cluster (Oracle Cloud Infrastructure Container Engine for Kubernetes) deployed on 51 standard 16-core virtual machine shapes (VM.Standard.B1.16). The revenue management layer persistence database was deployed in a full rack Exadata X7-2 hosting four RAC nodes. The details of the infrastructure setup are summarized in table 1.

	PROCESSOR	SHAPE	QUANTITY	NOTES
Test Client	X6	VM.Standard.B1.8 8 cores, 96 GB RAM, 4.8 Gbps	2	Intel(R) Xeon(R) CPU E5-2699C v4 @ 2.20GHz Total: 16 cores, 192 GB RAM
CCS Application Under Test	X6	VM.Standard.B1.16 16 cores, 192 GB RAM, 9.6 Gbps	51	Intel(R) Xeon(R) CPU E5-2699C v4 @ 2.20GHz Total: 816 cores, 9792 GB RAM
Persistence Database	X7-2	Exadata X7-2	1	4 RAC nodes Total: 184 cores, 2880 GB RAM

Table 1 – Scalability and Performance test cloud infrastructure details

Test Scenario

- 50 million subscribers with flat accounts
- Mixed traffic types across voice, data and messaging
- Mixed Charging Operation Types (initiate, update, terminate)
- 90% prepaid / 10 % postpaid per schema (4 schemas in total)
- A range of price plans covering prepaid and postpaid voice, data and messaging, including in bundle and out of bundle tariffs, friends and family rates

Observed Results

The Oracle CCS demonstrated single digit core charging latencies, high transaction throughput, efficient resource utilization and linear scalability.

Table 2 shows the transaction throughput and average observed latencies (all in single digit milliseconds), measured as a roundtrip between the network gateway internal charging requests and the core charging server instances.

TRAFFIC TYPE	TRANSACTIONS PER SECOND (TPS)	AVERAGE LATENCY (MILLISECONDS)
SMS	2,884	5.75
Voice	21,200	6
Data	76,800	6.96
Notifications	30,200	N/A
Balance Queries	5,592	2

Table 2 – Observed throughput and average latencies

Observed resource utilization across the core converged charging system application and the Oracle database used for persistence is shown in table 3, showing compelling results. Note that the utilization values shown in the table were observed during the steady state (maximum) traffic phase of the test.

RESOURCE UTILIZATION	
Average App CPU utilization	46%
Average DB CPU utilization	12%
Average App memory utilization	46%
Average DB memory utilization	54%
Max. App IOPS	300
Max. DB IOPS	7,360

Table 3 – Observed resource utilization

In addition, linear scaling characteristics were observed based on measurements from 12.5M and 50M provisioned test subscribers.

ALIGNED WITH THE INDUSTRY

The Oracle CCS is built around the standards of the industry, encompassing both the 3GPP charging standards and the TM Forum Frameworx and Open APIs initiatives. Oracle's product strategy is focused on constantly innovating and evolving a revenue management suite that helps to bridge the network and business domains by utilizing deep domain experience in data technology, communications business processes, core networks, and cloud technology.

Core Network Standards

The Oracle CCS supports concurrent charging for 4G and 5G data and communication services via the following industry specifications:

STANDARD	DESCRIPTION
3GPP TS 32.291 v15.4.0	Nchf Charging Service
3GPP TS 29.594 v15.4.0	Nchf Spending Limit Control Service
3GPP TS 29.510 v15.4.0	Interaction with the Network Repository Function (NRF)
RFC 3588	Diameter Base Protocol
RFC 4006	Diameter Credit-Control Application Protocol
3GPP TS 32.299, TS29.212	Support for 5G Non-Standalone Architecture (NSA) QoS-Information AVP extension for online charging
3GPP TS 32.299 v12.0	Diameter Gy Protocol
3GPP TS 29.219 v12.0	Diameter Sy Protocol

Table 4 – Oracle CCS coverage of 3GPP network charging standards

Figure 13 shows the Oracle CCS key network interfaces, with an example 5G data session flow shown in figure 14.

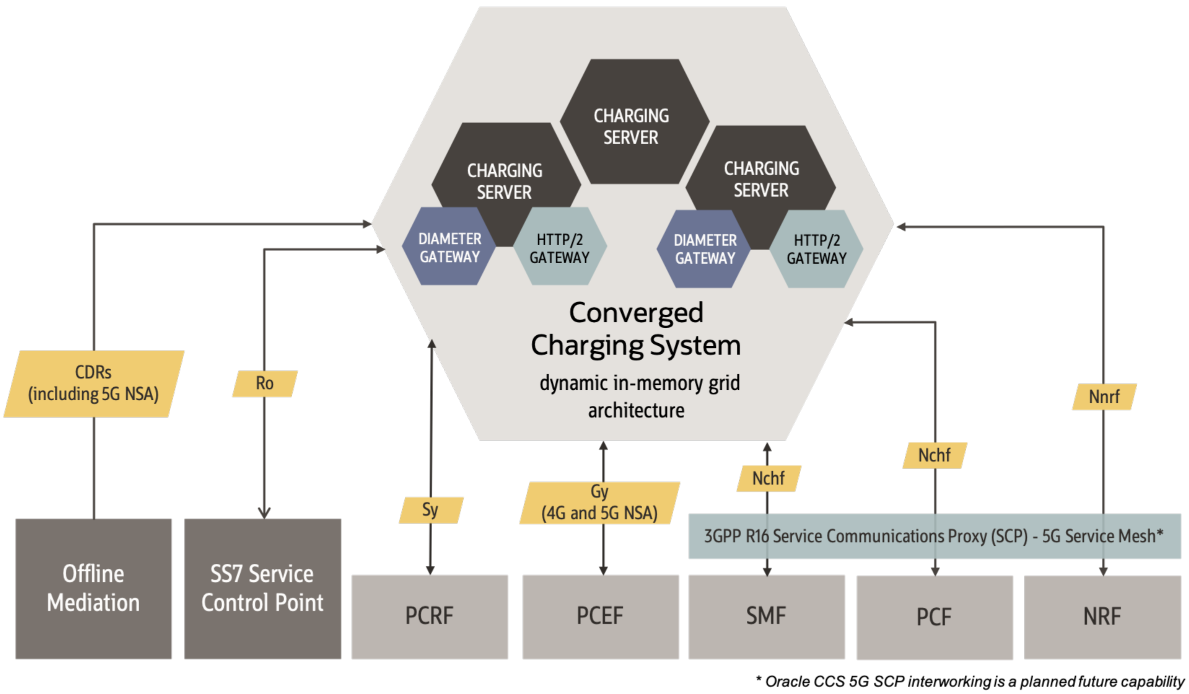


Figure 13 – Oracle CCS network context

In the 5G service-based architecture, the Service Communications Proxy (SCP) acts as a **5G service mesh**, which can be deployed to support traffic routing, load balancing, and prioritization across the 5G core (5GC) network functions. The SCP provides a similar function to the Diameter Signaling Router in a 4G context, both of which are network functions offered by Oracle. The SCP is a 3GPP release 16 function, with interworking support in the Oracle CCS a planned future capability.

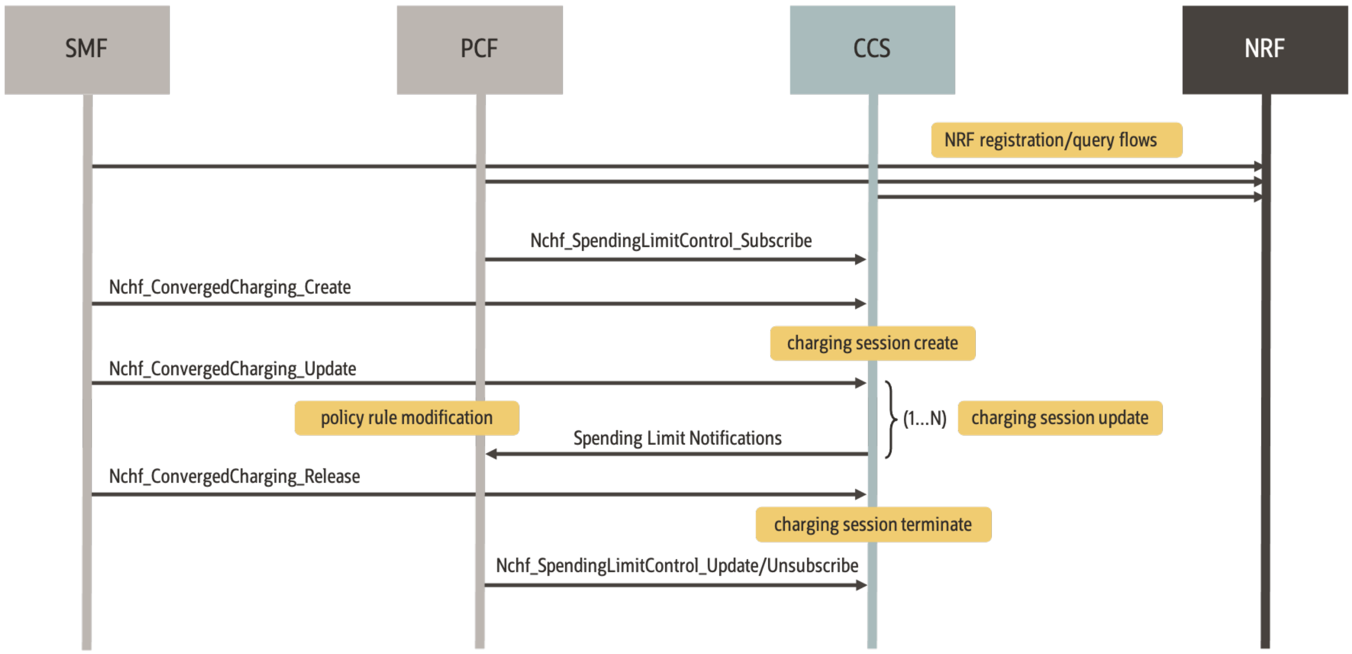


Figure 14 – Oracle CCS illustrative 5G data session flow

Integrated Policy and Charging

The Oracle CCS is integrated with the Oracle cloud native core Converged Policy solution, enabling 4G and 5G support for compelling dynamic policy enabled data plans based on factors such as bandwidth, QoS, usage volume, time of day, location and sponsored services. Figure 15 illustrates the solution functional coverage, adapted from 3GPP TS 32.240.

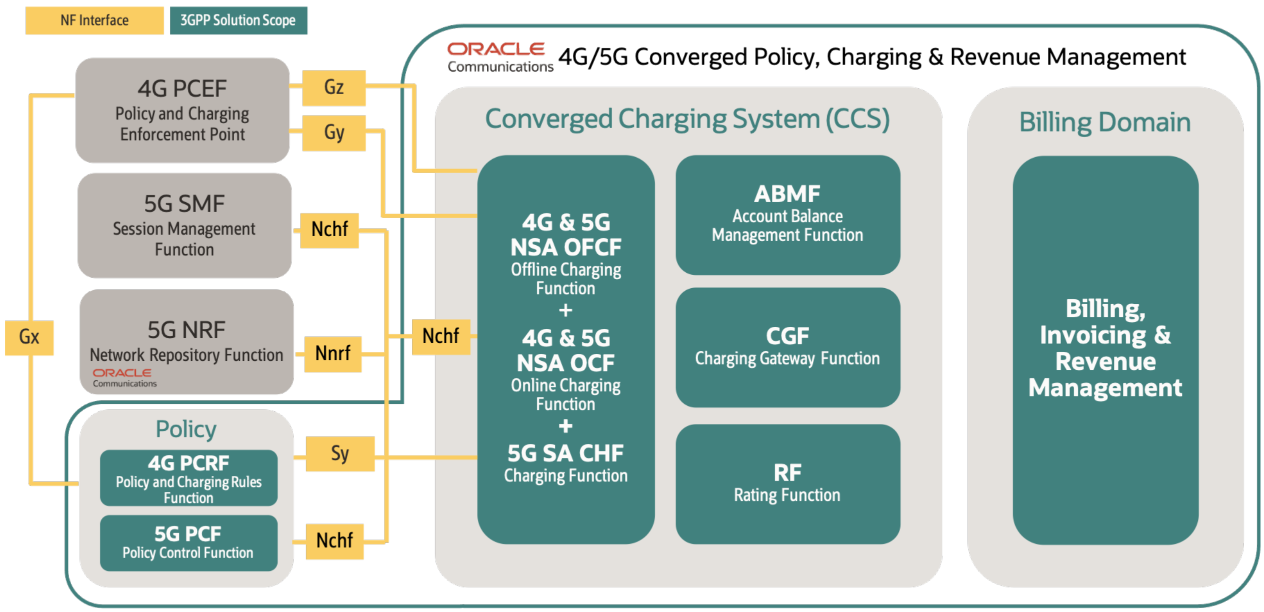


Figure 15 – Oracle Communications Policy, Charging and Revenue Management – 3GPP functional coverage

Figure 16 highlights the key solution components, including the capability for both policy and charging to be deployed on the Oracle Communications Cloud Native Environment.

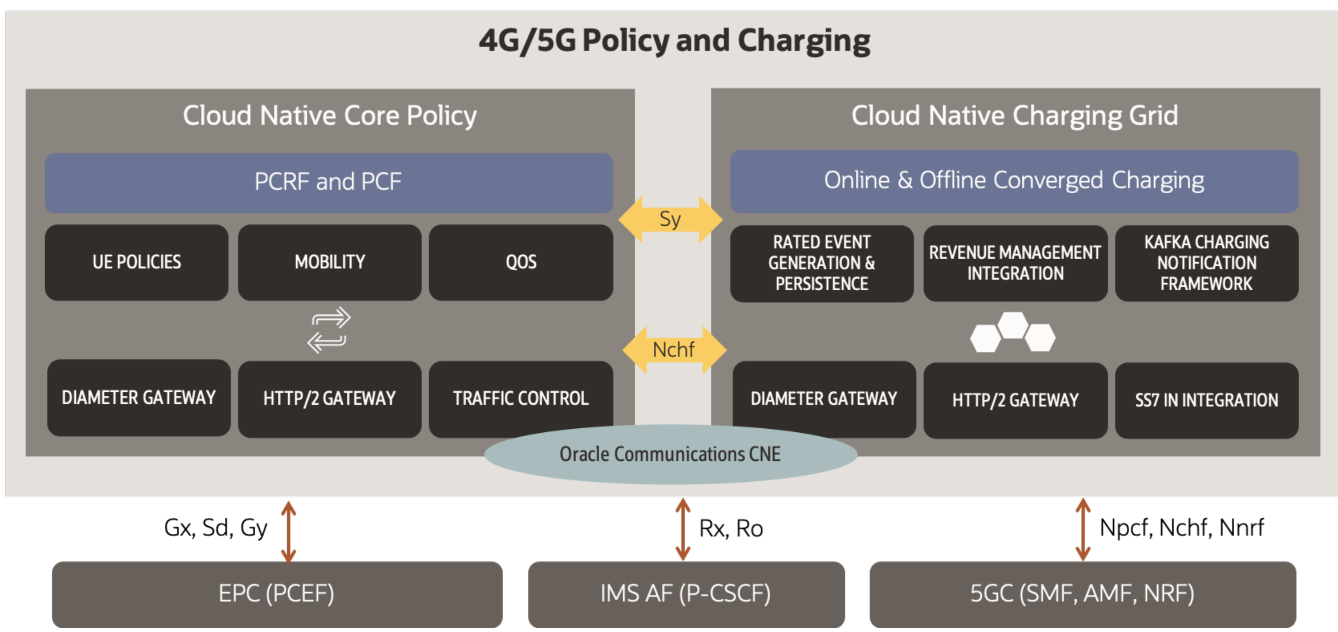


Figure 16 – Oracle Communications Policy and Charging – key solution components

TM Forum Open APIs

The Oracle CCS exposes a REST API for balance and consumption counters (based on TM Forum Open API 677) supporting various external services such as device-based self-care integration. Additional TM Forum Open API aligned REST end points are supported covering areas across product launch, buying and care.

CLOSE TO THE EDGE: A FOUNDATION FOR CHARGING IN THE 5G ERA

As sensors, device innovation, and network slicing drive the proliferation of edge-based computing, in which compute power becomes highly distributed in proximity to the Radio Access Network (RAN), the next generation of charging system will require the scalability, performance, and operational automation required to commercialize these services.

Edge computing, combined with the power of 5G network slicing, will impact both consumer and enterprise services, delivering rich, immersive content and supporting latency critical services (figure 17).

Figure 17 – Edge computing will impact all aspects of consumer and business life

Rich consumer-focused digital lifestyle offers that are monetized via the converged charging model will require millisecond response times to assure service quality and user experience. An example B2C use case is prepaid cloud gaming. A similar B2B/B2B2X example, industrial IoT, may additionally need to provide real-time dashboards about factory floor sensor consumption.

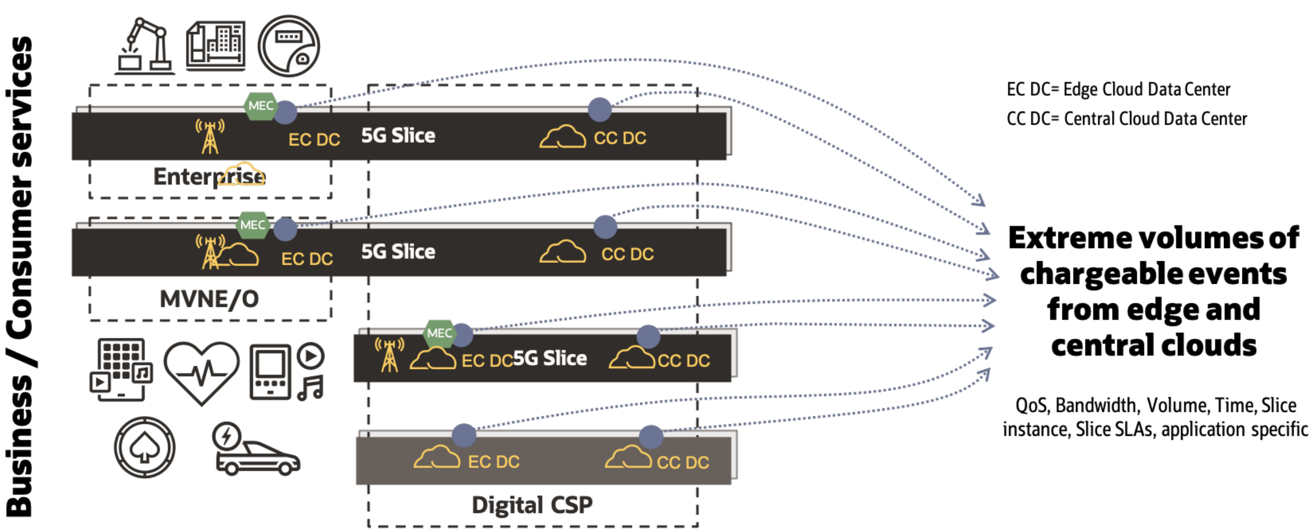


Figure 18 - Edge Computing will drive extreme volumes of chargeable events

Regardless of the path that edge computing and associated services will take, several things at least are clear:

- Extreme volumes of charging transactions will be generated from network slices serving enterprise, consumer, and B2B2X domains.
- The success of edge computing will be dependent on widely available cloud compute infrastructure.
- Edge based charging systems will require a multi-service cloud native architecture, aligned with industry standards such as the TM Forum Open Digital Architecture (ODA) and the 3GPP.
- The need for highly accurate, transactionally consistent charging for multiple concurrent sessions and complex account hierarchies and device ownership structures will be critical.
- Charging systems will require high degrees of operational automation.

The Oracle CCS is built to support a strategic vision that aligns with industry standards (3GPP and TM Forum Open Digital Architecture) and builds on accepted cloud native tooling and Oracle technology innovation (figure 19).

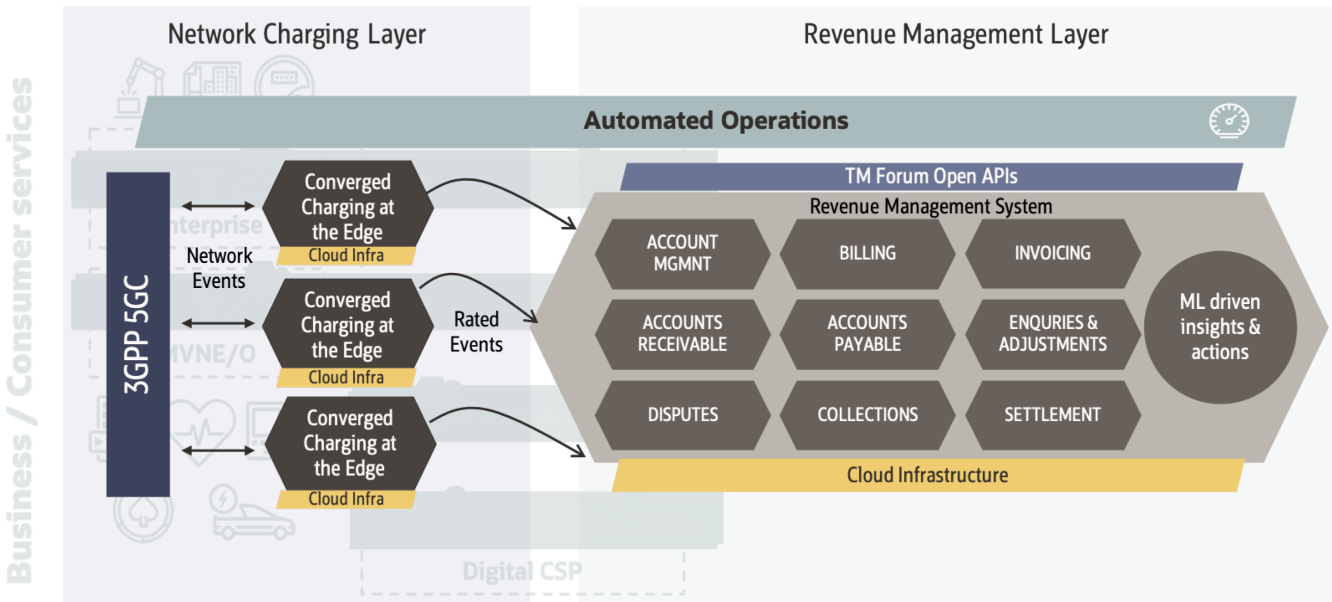


Figure 19 – Oracle CCS strategic vision – aligned with the industry, ready for the edge

SUMMARY

The Oracle Converged Charging System (figure 21) provides the following features to help communications service providers prepare for 5G monetization demands:

- Highly configurable, sophisticated, online and offline converged charging for any service, segment, and payment model
- A real time transactional system of record for converged data session charging and balance management, 3GPP aligned with native integration into Oracle's full suite of TM Forum aligned billing and revenue management applications
- Advanced 4G and 5G data session charging with adaptive quota allocation to reduce core network loads
- Elastic charging grid harnessing Oracle Coherence to enable extreme performance, elastic scalability, and cutting-edge resiliency across charging clusters, processes, and datastores
- Compliant with 3GPP Release 15 5G converged charging architecture
- 5G CHF/CCS and 4G Online Charging System (OCS)
- Supports both standalone (SA) and non-standalone (NSA) 5G core networks
- HTTP/2 gateway (Nchf and Nnrf interfaces)
- Diameter gateway (Gy and Sy interface)
- SIGTRAN-CAMEL (circuit-switched voice and SMS) supported through Oracle's Convergent Charging Controller
- Kafka notification framework
- Deployable with Oracle Communications Converged Policy to deliver dynamic policy enabled data plans

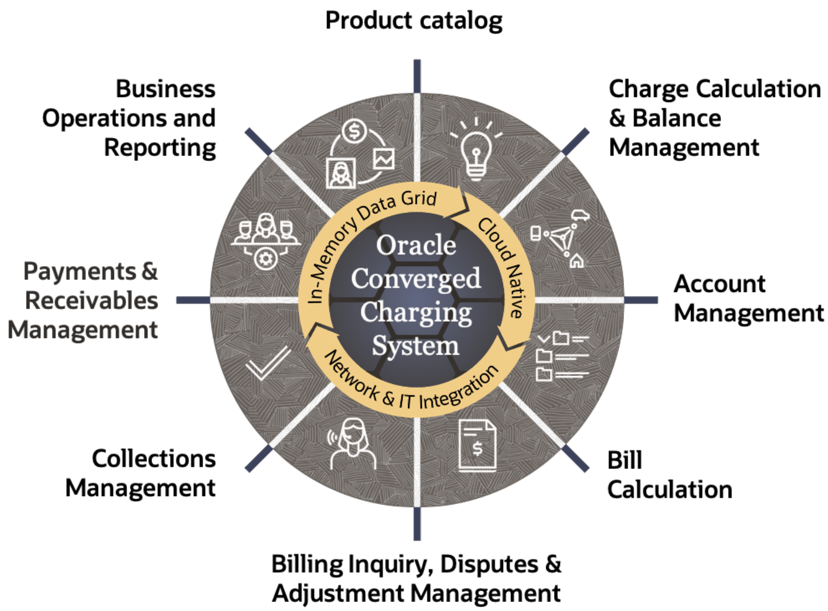


Figure 21 –Converged Charging from Oracle

The innovative in-memory mesh-based architecture provides a webscale charging experience with the following business benefits:

- Designed from the outset to support the technical and business monetization demands for hyperscale 5G digital communications providers
- Industry leading grid technology enabling high performance, low latency data and communications session charging and control with accurate rating and balance management
- Extreme transaction processing, almost zero cost locking with transactional consistency preventing the introduction of revenue leakage in the charging layer
- In-memory charging grid (not just a front-end cache), supporting collocated processing and data for massive parallel processing and cost-effective utilization of compute resources
- Near linear scalability due to the automatic partitioning of customer data objects across the grid members. A charging system to scale as your business grows.

- Fault tolerant state and data management to provide business continuity. Session continuity across failovers using Coherence Federated Caching supporting multi-site active-active deployments for the highest levels of availability, supporting a premium customer experience.
- Cloud native, ready to take advantage of modern compute, network and storage infrastructure with a self-healing containerized deployment architecture
- A future proof technical foundation for monetizing the world of rich 5G services delivered at the network edge with robust foundations for decentralized edge charging

GLOSSARY OF TERMS

5G NSA	5G Non-Standalone Architecture – deploying 5G New Radio (NR) with an existing 4G Evolved Packet Core (EPC). The most common starting deployment architecture for 5G.
5G SA	5G Standalone Architecture – the end game for 5G. A new service based 5G core (5GC) network using RESTful services and supporting network slicing (see below).
5G Slice	A virtualized partition across a 5G core network designed to support specific business segments with different technical characteristics, such as QoS, latency, and performance.
Active-active deployment	A disaster recovery (DR) architecture ensuring maximum business continuity in the event of unexpected site failure. Supported by the Oracle CCS.
Asynchronous persistence	Writing data to a persistent store in which the process requesting the write does not wait on a response (that is, it is non-blocking). Important to ensure minimal impact to real-time charging latencies.
CAP	CAMEL Application Part – the predominant Intelligent Network protocol. See also the entry for CAMEL below.
CAMEL	Customized Applications for Mobile network Enhanced Logic – the dominant IN architecture used to support circuit switched prepaid voice and messaging, including roaming.
CCS	Converged Charging System – a charging system that supports online and offline charging in a single architecture.
CDR	Call Detail Record – produced by core network elements during or after events have happened (for voice, data, and messaging services). In the 5G SA core, this responsibility shifts to the CCS.
Coherence	Oracle's industry leading in-memory data grid technology, powering the Oracle CCS.
Continuous Query Caching	An Oracle Coherence cache mechanism that combines a query result with a continuous stream of related events to maintain an up-to-date query result in a real-time fashion.
Converged charging	A concept that was first conceived in the late 1990s/early 2000s where a single charging system could support both online and offline charging. Previously referred to as "Convergent Charging", now adopted as a part of the 3GPP Release 15 5G standards.
Data affinity	A powerful Coherence concept where the data and processing acting on that data are co-located. This eliminates the transfer of data to the processing, supporting high degrees of parallel processing.
Diameter	The core charging and signaling protocol for 4G networks (twice the RADIUS).
Distributed Cache	A Coherence cache model in which data is partitioned across an entire cluster in a fault tolerant manner. Enables linear scalability in the Oracle CCS.
Federated Caching	An Oracle Coherence mechanism of replicating cache contents across a physical WAN. Used in the Oracle CCS multi-site active-active deployment model for maximum business continuity.
GGSN	Gateway GPRS Support Node (another nested acronym). A 2.5G and 3G node for supporting mobile data connectivity to a packet data network that generates data for processing by a charging system.
Helm	A package and deployment manager for Kubernetes applications.
HTTP/2	A revision of the Hyper Text Transfer Protocol (HTTP) supporting new features principally related to performance. It is used in the 5G service-based architecture to enable communications between the network functions.
IN	Intelligent Network – the concept of separating the service logic from the switching logic. The dominant mechanism for supporting prepaid circuit switched calls (for example, up to non-VoLTE 4G services) and messaging.
In-memory	The principle of storing critical data in fast memory so that no physical disk access is required for reads and writes (for example, directly from a relational database). Essential to support millisecond latency for converged charging operations.
Kubernetes	The dominant cloud native container orchestration tool.

MEC	Mobile Edge Computing – the concept of moving processing closer to the end consumer of the service, typically near the access network (for example, the Radio Access Network (RAN) in mobile networks). Likely to play a significant role in the future of digital communications and service delivery to support rich content delivery and ultra-low latency services.
Mesh	An architecture in which multiple nodes or functions co-operate to achieve a single goal. Examples include the 5G SCP (see below), service mesh technology such as Istio (used to support traffic management, security, and observability) and the Oracle Coherence in-memory grid technology (powering the Oracle CCS).
MSC	Mobile Switching Center – the 2G/3G mobile switching function. Deployed with a software function to trigger IN CAP requests for mobile prepaid where it acts as a Service Switching Point (SSP).
Offline charging	A charging model in which events are guided and rated after the event has happened. Typically used for post-paid charging and billing.
Online charging	A charging model which uses pre-delivery authorization with the core network to support real-time service delivery. Typically used for prepaid charging, but now important for converged charging across all end consumers for real-time spend control.
PCC	The 3GPP Policy and Charging Control architecture.
Pod	The Kubernetes deployment unit.
SCP	Service Control Point – an Intelligent Network function that is responsible for executing circuit switched SS7 based service logic. Oracle's Converged Charging Controller, integrated with the Oracle CCS is an SCP.
	Service Communications Proxy – a 3GPP Release 16 5G service mesh that is used to control HTTP/2 service request traffic in a 5G service-based architecture. Oracle was heavily involved in the standards development of the SCP network function.
REST	Representational State Transfer – a software architectural pattern for modern service API development, used both by 3GPP core 5G networks and also by the TM Forum Open API initiative. Supported by the Oracle CCS and integrated revenue management applications.
STDOUT	"Standard Output" – a UNIX/Linux file descriptor where processes can write output to. Used by the Oracle CCS containers to write log files for collection and visualization by Fluentd, ElasticSearch, and Kibana.
TM Forum SID	The TM Forum Information Framework, formerly known as the Shared Information Datamodel. A leading communications industry information model that the Oracle CCS bases its pricing and domain data model on.

CONNECT WITH US

Call +1.800.ORACLE1 or visit oracle.com.
Outside North America, find your local office at oracle.com/contact.

blogs.oracle.com facebook.com/oracle twitter.com/oracle

Converged Charging for Hyperscale Service Providers – Version 1.00
March, 2021
Authors: Richard Hallett