Edge computing and deployment strategies for communication service providers

Introduction

Communication Service Providers (CSP) are looking for new revenue sources to grow their businesses, especially in the enterprise area which will be increasingly important in the future. According to the Ericsson report "5G for business: a 2030 market compass", by 2030 up to USD 700 billion of the 5G-enabled, business-to-business value could be addressed by CSPs.

With the introduction of 5G and edge computing, they are now in a better position to provide new offerings both to enterprises that need to automate industrial processes and to consumers who require improved user experiences for on-line gaming. Edge computing provides distributed computing and storage resources closer to the location where it is needed and targets new business opportunities that provide support for specific application use cases. Some examples of use case areas are augmented and virtual reality, manufacturing and automotive. The innovation rate in this part of the application ecosystem will be significant going forward.

The edge opportunity should be seen in a larger context of the enterprise opportunity, where edge computing will be an enabler for many broader use cases, for example within the Internet of Things (IoT) and potentially bundled with other enterprise offerings such as 5G private networks.

The ecosystem for edge computing is fragmented and is quickly evolving. Technical solutions - interfaces, standards and business models are not set. Several players must be involved to create end to end solutions and CSPs must carefully consider in which industries they can expand their offerings in beyond connectivity.

The edge application ecosystem is driven by third party applications outside of the telecom domain since solutions for new use cases require specific domain knowledge from industry players outside the telecom space. Edge infrastructure will therefore be accessible to third party application providers and developers and will host a multitude of applications, each with specific characteristics and needs.

This white paper describes edge computing from a CSP perspective, - how a solution is built up, the key industry challenges and how CSPs can choose different roles and deployment strategies when addressing opportunities. Depending on their market position and type of use case provided, they can choose an optimal model or a combination of models that suit their needs.

Functional components of an edge computing solution

The edge application environment enables CSPs to host non-telco workloads and open up the network as a distributed cloud resource. Enterprises can develop applications, deploy and manage them flexibly via orchestration logic towards a landing-zone which accesses the distributed cloud infrastructure and leverages services exposed through APIs for consumption. Below is a brief overview of the functional components needed to create an edge computing solution.

Distributed cloud infrastructure

This is a combination of different sizes of cloud data centers at global, national, local/regional and potentially access locations integrated to the network and operated by a central orchestration and management system. The exact specification of the infrastructure on the different sites may depend on the use cases and applications onboarded. In addition, there can be several infrastructure providers in the same site.

Connectivity

Once the development environment is installed, connectivity shall be configured. It might be specified by the application developer. The application running on the network edge may have connectivity requirements on bandwidth, throughput, mobility and/or latency within its components (for example, deployed on different hardware in a redundant setup) or with the external world, such as an internet connection, and the user equipment or the application session. Traffic routing for applications deployed further out in the network topology will need new mobile network solutions, such as distributed anchor, session breakout and multiple sessions, and in some cases coordination between application server selection and usage of these mobile network solutions.

Application runtime execution environment

The very basic functionality that an edge computing service may provide is the runtime execution environment (RTE) for virtual network functions (VNF) and non-telco workloads. An execution environment should be able to host applications and harmonize the requirements of the development communities.

Many applications may use edge computing with different characteristics and functional requirements and will then require different platform components. Therefore, the operator provides a generic or multiple execution environments on the network edge that can be later customized by application developers.

Dynamic orchestration and management

There should be a central orchestration and management functionality that is aware of the network topology and what resources are available where in the distributed cloud infrastructure. In order to keep consistency between possible traffic breakout points (where user plane gateway functionality is deployed) and the applications (which consume the traffic) in the network edge. This orchestration layer will provide a harmonized single orchestration and management functionality over the different orchestration functions that will be present. It is to be used to manage the platforms for non-telco workloads and VNFs according to service level agreements.

Service exposure

Exposure is a key function to define and develop new capabilities (APIs) and securely expose them to non-telco workloads. The exposure server exposes the core capabilities available internally within the operator or to a partner with whom there is a commercial agreement. The exposed core capabilities add value to internal or external users, for example, connectivity, optimization, identity, security, data and analytics.
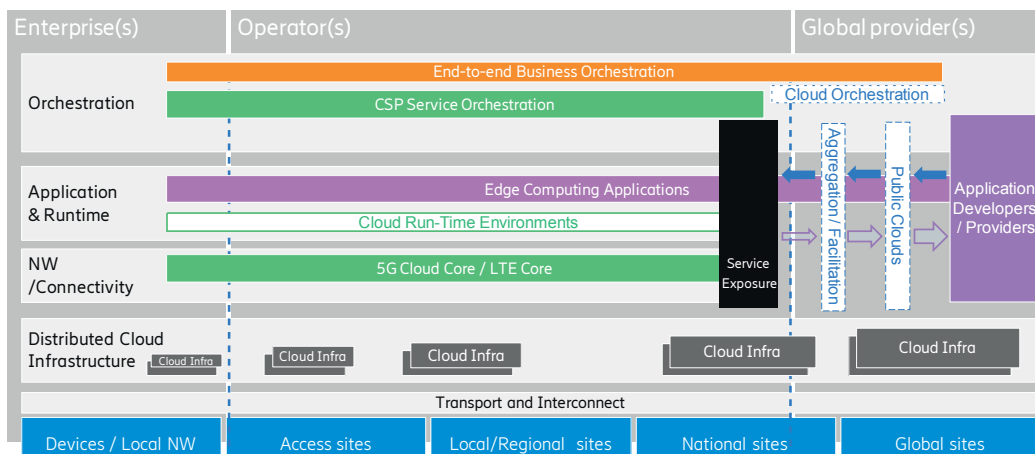
The edge computing landscape

A vast number of companies participate in the ecosystem, from hardware vendors, platform companies to applications developers, System Integrator (SI) companies and CSPs.
Two other key players in the ecosystem are the Hyperscale Cloud Providers (HCP) and Operations Technology (OT) vendors.

HCPs, such as AWS, Microsoft Azure, Google and AliCloud have the core business to provide cloud infrastructure and platforms. They own application ecosystems with thousands of contributing developers and can serve multiple enterprises in several sectors on a global basis. HCPs are keen to be ecosystem drivers for edge computing.

OT vendors have IoT platforms and applications, supported by edge computing components. Some examples of these companies include Siemens, General Electric, BMW and ABB. They have strong enterprise relationships, especially in the manufacturing sector. Companies looking to do a smart manufacturing deployment is likely to partner with an OT vendor to a certain extent. OT vendors are establishing relationships with HCPs for global deployments of their solutions, access to application development ecosystem and an environment to create and deploy their own applications.

SI companies have a wide range of capabilities to address enterprise pain points related to solution implementation and integration of offerings from different ecosystem companies. SI companies can be both global and local and are likely to be present in most solution implementations in one way or another. Apart from specialized SI companies, other companies can also take an SI role in a solution implementation, for example OT vendors or HCPs.

Edge value stack layer definition

From a market and ecosystem perspective, the functional components described above can be translated into a value stack where each layer in the stack is addressed by competing companies.

Services	– Consulting, systems integration for different layers (connectivity, applications, etc.) and solution lifecycle management
Application development	– Software running on top of the platform and consuming its services, either for industrial or consumer use
Application deployment and enablement platform	– Platform as a Service layer to enable deployment of applications and provisioning – Generic application services (exposure parts and implementation of exposure services, for example connectivity exposure services)
Software infrastructure	– Software required to provide infrastructure, virtualization layer and Containers as a Service – Software for hardware management and orchestration to support execution of platform orchestration commands
Hardware infrastructure	– Hardware infrastructure (servers, routers, storage, cabling, racks, etc.)
Connectivity	– Connectivity - fixed or mobile providing IP presence for edge deployments
Site	– Where hardware infrastructure is placed, such as CSP premises, co-location companies, tower companies or the enterprise's premises

Even though the value stack looks uncomplicated, a vast number of companies participate in the ecosystem to take a role and it can therefore be challenging to navigate in it. Many companies also have the capability to address more than one layer.

Edge computing is specified across several standards and open source fora

The edge computing ecosystem is very dynamic with many new initiatives from various organizations and companies. Just in the open source area there are at least 20 initiatives ongoing across different communities as this paper is being written. There is no industry standard agreed as yet covering all aspects of edge computing, even after years of efforts by some standardization bodies. Edge computing is being specified across standards and open source fora, for example 3GPP which has accelerated its activities towards edge, ETSI, CNCF (Cloud Native Computing Foundation), ONAP (Open Network Automation Platform) and LF Edge. In addition, there are several industry alliances aligning around their own use cases. Two good examples are AECC (Automotive Edge Computing Consortium) and 5G-ACIA (5G Alliance for Connected Industries and Automation).

Standardization bodies	Open source foras	Industry alliances
ETSI	CNCF	5GAA
TM Forum	LF Edge	AECC
3GPP	ONAP	5GACIA
	OpenStack	Industrial Internet Consortium
	Edge Computing Group	

Avoiding ecosystem fragmentation

The first challenge for CSPs and the industry as a whole is to avoid ecosystem fragmentation. When many organizations work to specify how edge computing should be deployed, there is an obvious risk for fragmentation, which means that standards, technology, interfaces and business models do not match leading to slower uptake of new services and not reaching the economy of scale required.
Standardization will be key on low level technical APIs, but for applications and exposure on the BSS (Business Support Systems) layer which, will allow securitization and monetization of the APIs, higher level APIs are needed. Those high-level APIs will at least initially be de-facto standard defined as part of the implementation and will not be coming from standard bodies even though they may end up there eventually. The industry needs to avoid fragmentation but at the same time needs to allow for differentiation and competition between CSPs.

The industry can use the following standards and specifications on the technical level already now as a base for edge computing implementations.

— Distributed cloud infrastructure manages hardware and infrastructure services. Multiple layers of distribution allow for a flexible definition of 'edge'. ETSI-NFV should define the required interfaces.

— Orchestration and management of workload services on top of a distributed cloud infrastructure is done by dynamic orchestration, embracing multi-domain orchestration. TMF (Telecom Management Forum), ONAP, 3GPP and ETSI-NFV (MANO parts) are most suited to define the relevant interfaces and functions.

— Connectivity and mobility handling functions and interaction with communication networks, such as 4G and 5G, define the possibilities to connect mobile subscribers to services inside or outside the CSP domain. The relevant functions and interfaces are defined by 3GPP.

— The development environment is foundational for the rapid creation of new services. By exposing cloud, network and orchestration resources, software developers can improve their edge applications with various functions from the CSP. Application developers expect that the telecom networks are abstracted so they can avoid that complexity and focus on the application logic. They should not have to understand network topology, legacy protocols and APIs. Instead, standardized and cross-industry aligned APIs are needed. CNCF builds sustainable ecosystems and fosters communities to support the growth and health of cloud-native open source software. Furthermore, Kubernetes provides tools and components to help developers build, deploy and execute cloud-native applications.

Partnering to grow the opportunity

Edge computing should be seen in the context of the wider enterprise opportunity and as an enabler for many broader use cases, for example within IoT, and potentially be bundled with other enterprise offerings such as private networks. This means that CSPs must have an enterprise strategy to start with. The strategy takes into account the enterprise landscape in the region and the capabilities beyond connectivity that can be provided to enterprises.

A key factor to succeed in delivering new services is to have a strong business relationship with the enterprise customer. However, these customers may not think of the traditional CSP as a natural provider of solutions, for example to automate processes in a factory. Overcoming this challenge will require the adoption of go-to-market (GTM) strategies that allow selling edge solutions that meet the different use-case requirements in various industry verticals.

CSPs need access to the adequate domain knowledge through partners, such as SI companies, HCPs and the OT vendors, as well as increasing credibility through association with the right brands and industry experts as the other players will need the network and connectivity knowledge that CSPs provide. Partnering with these companies will in many cases be necessary to reach the application ecosystem to enable use cases and play a relevant role in providing those.

The need for differentiation

How can CSPs leverage a global ecosystem for edge computing and collaborate with these companies while differentiating locally?

Service exposure is an important enabler to create a good developer experience, to create value and provides an opportunity to differentiate oneself. CSPs should work on differentiating themselves across the services, capabilities and characteristics they offer through easy-to-consume APIs, but not in that way the services are described or consumed, i.e. not on the APIs themselves. It is in their interest to align behind a common way of interacting with the connectivity services. A good example of an exposure success story is the HCPs that have attracted application developers on a global scale through simple to use APIs.

Below follows a three-step approach that can be used when exposing services through APIs, to enable a global ecosystem of applications while still offering differentiating services.
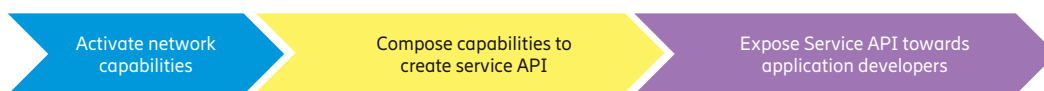
Activate network capabilities	Compose capabilities to create service API	Expose Service API towards application developers
Based on Standards	De-facto adoption and open source and important play	

In the first step, 'Activate network capabilities', standard network capabilities are activated and exposed. This means that those network capabilities are made available in the network but also exposed for consumption by external entities. The technical APIs are defined by standards from 3GPP and TMF for example.

In the second step, 'Compose capabilities to create service API', edge use cases are supported by the composition of technical APIs (network and management) enabling the creation of service APIs tailored around specific use cases. Service APIs will be adopted de-facto when used and proved valuable for an edge use case. The service exposure here covers:

— Standardized slice characteristics for pre-integration: latency, bandwidth, proximity, QoS, rating.
— Value added services for differentiation: security, geo fencing, etc.
— Instrumentation: insights and monitoring.

Just building a set of APIs that make sense and address specific use cases will not be sufficient. Those APIs must also be made available to application developers during development time to enable them to incorporate them in their work. This is done through Software Development Kits (SDKs) that can be integrated in development environments, including those provided by HCPs. So, in the third step, 'Expose service API towards application developers', the service- and instrumentation APIs defined in the second step shall be the foundation of the integration with HCP and OT vendor developer environments. These APIs secure pricing/monetization, lifecycle management and security mechanisms.

The means of integration to application development environments will be done based on an SDK and a set of APIs. There are challenges that must be addressed by that integration. Firstly, on development time already covered by step two with de-facto standard APIs described above. And secondly, on execution time by addressing the discovery of the services exposed but also the monetization of the APIs in step three.

Deployment strategies

Based on their enterprise strategies, the use cases addressed and the underlying business case, CSPs can take different roles in the value chain. The roles are categorized based on whether they want to build edge infrastructure and whether they want to front the enterprise. Fronting the enterprise means that CSPs have more than the relationship for connectivity, but also the relationship to influence the enterprise choice of edge deployment setup. One can take a single or a combination of the roles described below as part of the strategy, for example towards different industry verticals depending on how strong their relations.

Full Edge Provider

— The Full Edge Providers have a strong go-to-market (GTM) relationship with the enterprise or the application developer/provider. Could be in partnership with for example OT vendors.
— They also provide edge infrastructure and potentially platform. Could be separate or the same infrastructure used for telco workloads. The Full Edge Provider commits to SLAs.

Partner Edge Provider

— Partner Edge Providers have a strong GTM relationship with the enterprise, especially for edge use cases strongly linked to connectivity.
— They provide connectivity, resell HCP and OT vendor infrastructure and platform and can host their edge stack. The Partner Edge Provider commits to the SLAs.

Aggregator Edge Provider

— A Content Aggregator provides infrastructure software and deployment platform as-a-service and commits to SLAs towards an application developer with strong GTM towards enterprises. The Content Aggregator could also have direct GTM towards the enterprise.
— The Aggregator Edge Provider (the CSP) has the edge hardware, separate or the same infrastructure as for telco workloads.
— Limited Edge Provider
— HCP, OT vendors and SI companies have horizontal and industry vertical capabilities and strong GTM. They front enterprises for most of their needs, including edge infrastructure and platform, and commit to SLAs.
— The CSP scope is focused on connectivity and, in some cases, co-location. The connectivity offering can be extended depending on who fronts the enterprise.

Limited Edge Provider

— HCP, OT vendors and SI companies have horizontal and industry vertical capabilities and strong GTM. They front enterprises for most of their needs, including edge infrastructure and platform, and commit to SLAs.
— The CSP scope is focused on connectivity and, in some cases, co-location. The connectivity offering can be extended depending on who fronts the enterprise.

Choosing a role in the value chain

How does a CSP know which role to take of the four described above? In order to analyze that and select a suitable deployment strategy one can go through a set of questions addressing;
— Level of enterprise growth ambitions and willingness to invest
— Market size and position in that market
— Level of current enterprise relations and GTM capabilities
— Competence in enterprise verticals
— Capacity and competence to provide complex solutions

CSPs have different ambition levels for enterprise customers. Some only focus on connectivity while others want to go beyond connectivity and also provide complete edge computing capabilities. In addition, they have different capabilities to scale and address a larger market. By combining these two factors, five different segments emerge:

Ambition in enterprise segment
(to front enterprises beyond connectivity)

High	Technology innovators with small scale	Local champions & mid-market ambitious	Enterprise giants
Low	Long tail	Connectivity-focused heavyweights	
	Low	Medium	High

Capability to scale & size of addressable market

Enterprise giants: Large CSPs with strong enterprise business in large markets. They see edge computing as a critical enabler for their enterprise strategy and are a in strong position to invest in GTM and technical solutions.

Local champions and mid-market ambitious: Leading CSPs in small/mid-size markets and mid-size in large markets, with high ambitions to grow their enterprise business beyond connectivity.

Technology innovators with small scale: Small or mid-size upstart CSPs with high technology ambitions and willingness to grow their business beyond connectivity.

Connectivity-focused heavyweights: Mid or large-size CSPs focused on connectivity offerings.

Long tail: Small size CSPs focused on connectivity offerings in local markets.

Recommendations for high ambition segments

How should the different segments address edge computing by using the deployment tracks described above? Depending on the industry vertical, position in the market and similar factor, CSPs can choose different roles, or a combination of roles in the value chain. The following provides some generic recommendations based on which segment they belong to.

Enterprise giants should invest in enterprise GTM to avoid commoditization (strive for Partner Edge Provider rather than Limited Edge Provider) and consider investing in own edge infrastructure (Full Edge Provider). They should focus on a handful of verticals and use cases that can scale and be replicated horizontally and drive innovation and standardization of new APIs in these verticals to maximize value from connectivity. In addition, they should form multiple partnerships and strive for consumption-based revenue models and for controlling parts of the higher layers of the stack such as orchestration, portal and exposure.

Local champions and mid-market ambitious should invest in enterprise GTM to avoid commoditization focusing on a couple of verticals where they are the most competitive in. They should focus on building partnerships whose offerings they can reuse and resell (Partner Edge Provider and Aggregator Edge Provider) and selectively get involved in opportunities to innovate and co-create together with enterprises and other partners.

Technology innovators with small scale should leverage multiple partners (Content Aggregators, OT vendors, HCPs, and SI companies) to rapidly gain scale (Partner Edge Provider/Limited Edge Provider or Aggregator Edge Provider) and invest in enterprise GTM where they are able to develop strong relationships with local customers and vendors.

Conclusion

With increasing interest in new use cases and services like smart manufacturing, augmented and virtual reality and the high interest in online gaming, there is a clear need for edge computing. However, the edge is not a standalone product or an offering but an enabler for use-cases requiring security, resilience, and low latency in combination with other technical solutions like private networks.

The edge computing ecosystem is vast and is evolving rapidly. Many organizations and companies are involved in specifying the technology and defining solutions. This runs the risk of market fragmentation leading to slower uptake of services. The industry can avoid fragmentation by making sure that differentiation is done on services, not on how they are consumed or exposed.

Edge computing covers a vast number of uses cases, but there's no one solution that fits them all. CSPs should choose the one that suits their enterprise strategy best, and should be prepared to build strength by partnering with HCPs, SI companies or OT vendors, while keeping a strong GTM towards enterprises.

CSPs can choose different roles, a single role or a combination of roles, and deployment tracks depending on their ambition in the enterprise area beyond providing connectivity. These roles are categorized based on whether service providers want to build edge infrastructure and whether they want to front the enterprise. Depending on the ambition within the enterprise segment and the potential to scale, five different segments emerge where the recommended strategy will vary between the segments.

Further reading

Edge computing
Learn more about Ericsson's edge computing approach from an end-to-end perspective.

Webinar with Mobile World Live
Watch the webinar "Harnessing Edge Computing for 5G Success".

Edge Gravity
Read about our Edge Cloud Platform that facilitates the global collaboration between content, application, and service providers to deliver services meeting customer's highly interactive and data intensive needs at the edge.

Contributors

Carlos Bravo, Director Cloud Strategy Execution, Madrid

Carlos works as Director Cloud Strategy Execution at the Ericsson CTO office, leading multi-disciplinary teams to develop and implement cloud strategic initiatives. Carlos has worked for Ericsson since 2000, from Service Delivery to Product Development, Global Services, Market Unit and Business Area, mainly in the area of OSS/BSS as Architect. He started the Cloud journey in 2010, delivering cloud solutions to customers and then being lead architect of the Ericsson cloud solution. Carlos holds a MSc. in Telecom Engineering Data Communications.

Henrik Bäckström, Solution Marketing Manager, Stockholm

Henrik works as Solution Marketing Manager at Business Area Digital Services and focuses on running marketing campaigns for Edge Computing and Cloud Infrastructure solutions. He has worked for Ericsson since 1999, starting with product management and commercial management before going into marketing with experience from access, core networks and IoT. He started to work with cloud products and NFV in 2014 responsible for several market launches. Henrik holds a MSc. in Business Administration.