2019

chetan sharma
CONSULTING

Edge Computing Framework:

Understanding the Opportunity Roadmap

Chetan Sharma

Table of Contents

Disclaimer

Chetan Sharma Consulting is one of the most trusted advisory firms in the global mobile industry. This research document presents some in-depth analysis about the future of the mobile industry. However, the author or the company assumes no liability whatsoever.

Any use or reprint of the material discussed in the paper without prior permission is strictly prohibited.

Introduction

In 1910, writing a newspaper column, the legendary inventor Nikola Tesla quipped, "The spread of civilization may be likened to that of fire. First, a feeble spark, next a flickering flame, then a mighty blaze, ever increasing in speed and power."[1] As has been evident from the past 50 years of technology evolution, the emerging notions of computing and communications follow the same script. With each new breakthrough, the impact of how computing and the power of communications is adopted grows bigger in magnitude. We are living through an amazing technology renaissance wherein several economy-altering trends are shaping our world at the same time.

The notion of industry disruption as new technologies are introduced is not new but each time it is a bit different. In the Connected Intelligence era, we are seeing accelerated pace of change like never before. All this is leading to the evolution of computing itself. From mainframe to client-server to cloud, computing has fundamentally altered the global industries, the way we communicate and interact with information.

Now, thanks to convergence of the forces mentioned above, we are in the early stages of a new form of computational model that takes advantage of the resources available in the connected world – Edge Computing. The compute platform can reside at any point of the end-to-end network to enable faster processing, creation of actionable knowledge, and enable new ways to distribute data and intelligence that were not possible before. In fact, Edge Computing will transform every aspect of the computing and communication ecosystem and the value chain.

[1] Source: What Science May Achieve This Year, Nikola Tesla, Denver Rocky Mountain News, Jan 16th, 1910

Edge doesn't replace cloud but rather supplements it by sharing the workload in scenarios that are not well-suited for the traditional cloud architecture. Additionally, some of the emerging use cases in multi-player gaming, AR/VR, autonomous vehicles, using facial recognition for security, content especially video management, offloading, and delivery.

We are still in the early stages of understanding the capabilities and the potential of Edge Computing. There are many use cases that are already in the works. We don't entirely know which horizontal and vertical segments will be the biggest winners or which specific featureset will be most used by the developers or how fast they will scale. Each entity will look at the Edge Computing opportunity from their own lens be it their vertical industry or a tool that serves the ecosystem. We know that Edge Computing will redefine how we deal with bits and bytes but what will be the contours and trajectory of this set of opportunities.

We think the opportunity is best understood with the help of the developer community and the ecosystem. To start an industry conversation around Edge Computing opportunity, this paper presents an Edge Computing Framework to look at the opportunity across multiple dimensions: horizontal areas, verticals, functionality, market size, location and time to market. We will try to understand the opportunity using some current use cases and present our first cut of the framework output. By refining it over time as more insights and data become available, we will be able to understand the long-lasting impact of Edge Computing and shape the opportunities to come.

Defining Edge Computing

While most observers agree that Edge Computing is a paradigm shift in computing, there is no singular definition of Edge Computing that has gained consensus. According to ETSI, "Mobile Edge Computing provides an IT service environment and cloud computing capabilities at the edge of the mobile network, within the Radio Access Network (RAN) and in close proximity to mobile subscribers." IEEE defines Edge Computing as something that places applications, data, and processing at the logical extremes of a network rather than centralizing them. Linux Foundation says something similar - "The delivery of computing capabilities to the logical extremes of a network in order to improve the performance, operating cost, and reliability of applications and services."

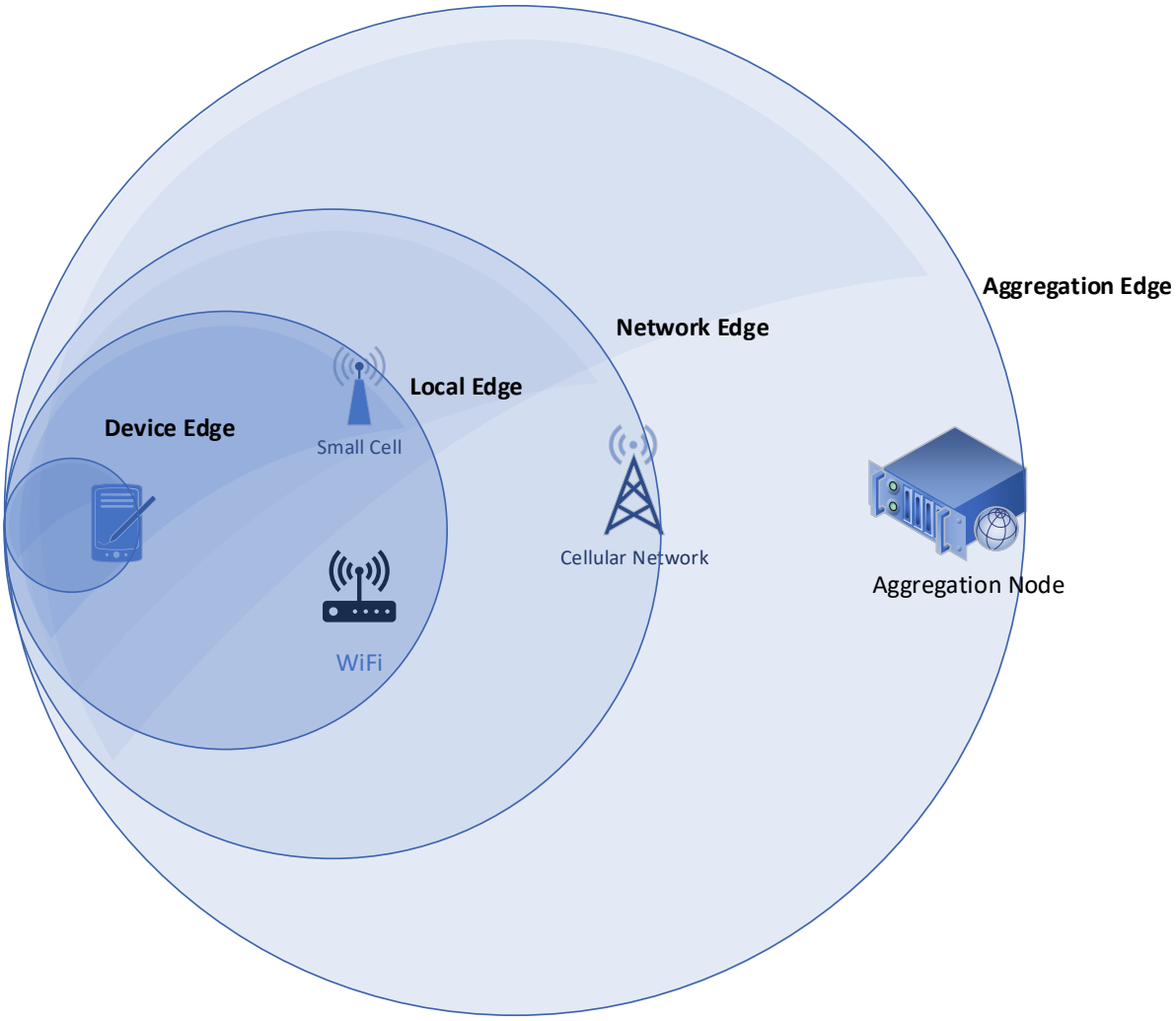

Figure 1. Edge Computing Network

But what are the "logical extremes" of the network? In most instances, they are the end-user devices – smartphones, sensors, automobiles, health monitors, etc. From an Edge Computing perspective, the location (not necessarily the physical location) is defined by the task at hand and how do you get capabilities to the node that needs it in a most cost-effective and scalable manner. For e.g.

- If one is looking for location data for a few hundred devices at the same instance in a local district, the base station is ideally most suited to have that information that can be dispersed to the entity that needs it, so the base station is the edge in this instance.
- For a drone that is collecting images in real-time on an oil pipeline and is trying to determine if there are noticeable changes, the drone itself is the edge device which can do this determination.
- In some instances, the vehicular traffic data must be aggregated to predicting congestion in cities. In such scenarios, the aggregation layer is the edge.
- In healthcare use cases, depending on the state of the application and the task supply-chain, the edge could vary from being at the device on the patient (like smart watch or the heart monitor), at the RAN (in operator's network), at the local hospital chain (that is processing the data from multiple patients) or at data aggregation point (that is looking at the national data streaming in).
- In the case of facial recognition based security systems at schools, using facial vectors for authentication can be done using a local small cell or access point (in which case that will be the edge) while if you are trying to compare a face to a wanted database, typically, you will do it with the help of aggregation nodes or centralized cloud as such the edge moves back in the network. However, if you are doing a recognition to validate the user for a device, the device itself is performing all the computations and authentication and becomes the edge.

The key in all these examples is that the intelligence is programmed for the highest efficiency. Resources are instantaneously distributed to the node closest to the point of need. This is done through the capabilities available at the device, network, and the cloud. It should be internalized that Edge is not optional but rather a core element of the framework through which every application and service will be designed. In the next section, we will take a look at the market forces that are enabling the Edge Computing cycle.

Market Forces Enabling Edge Computing

We are going through a unique phase of technology evolution when many symbiotic breakthroughs and generational changes are taking place at the same time. The wireless industry is kicking off its 5G cycle that is bringing in a new set of capabilities to the market such as enhanced broadband, lower latency, support for denser IoT deployments, network slicing, tighter security, and much more. The sensors are being miniaturized and becoming embedded into everyday things. The computing platform is available to these sensors for not only processing basic data emanating from the sensors but also to the AI algorithms that can make sense of the data in real-time. Finally, the new genre of applications and services are emerging that require a much more distributed approach to information distribution and processing.

Over the past few years, we have been making tremendous progress on several fronts:

Hardware: Sensors have become much cheaper and smaller to the point that we are putting a sensor in almost everything.

Data: The availability of sensors at the edge has meant that the amount of data available to derive intelligence has grown manifold.

Software: Finally, the industry has emerged out of the AI winters of the past to truly make AI a programmable component of the software stack in a way that can ingest data and automate functions in a rapid manner. The engine can self-correct and self-learn as more data becomes available.

Additionally, the nature of the wireless networks is changing. From relying primarily on the macro network of base stations across the country to now densifying the network with the help of small cells and access points (Figure 3). 5G will further accelerate densification of the network with the number of small cells more than double in the next five years. This means that first network touch point for the device will be much closer than it used to be. Smartphone itself has changed tremendously. The latest iPhone is capable of 5 teraflops which is more than what a cloud server was able to do only five years ago (Figure 3). Smartphone will have the capability of acting as an edge device to other smartphones for location, content, and data processing.

The number of connections is expected to grow tremendously over the next five years, especially on the IoT front. Compared to 2018, the number of IoT connections will more than double by 2024. Similarly, as we kick-off the 5G investment cycle, the number of 5G subs will exceed over a billion subscribers in the same time period according to GSMA. (Figure 4).

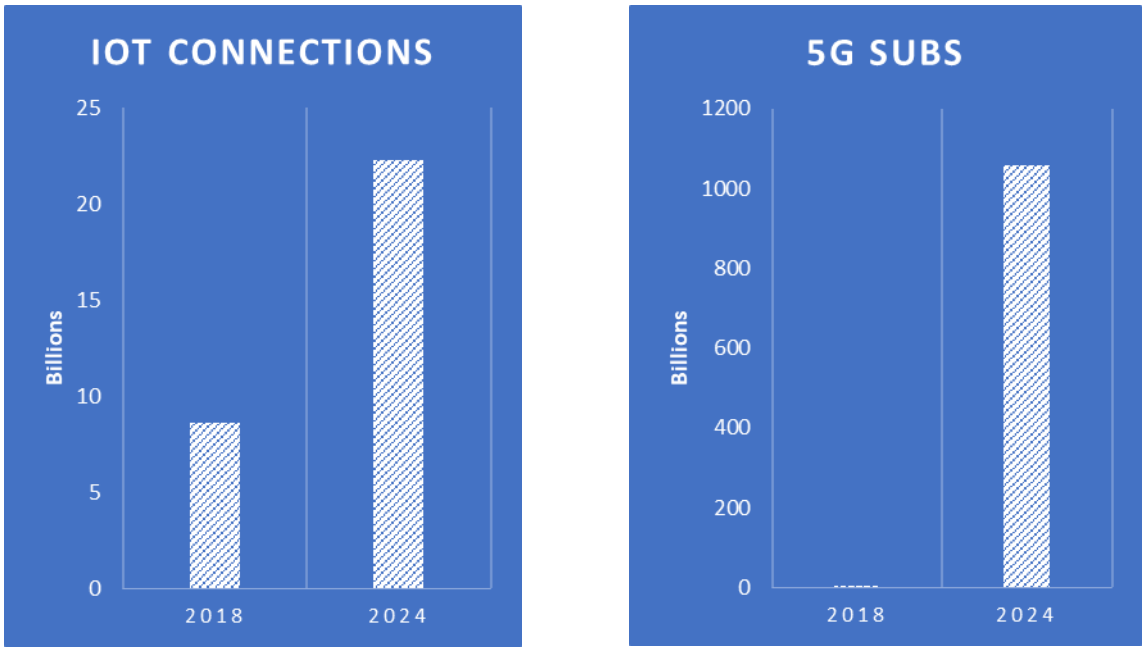

Figure 3. Growth in IoT Connections and 5G Subs[2]

One of the most significant challenges for the next decade will be how the networks manage data traffic. Mobile data traffic is exploding worldwide. India's Jio has been clocking 10 GB/sub/month for over 250M subscribers. Finland's DNA announced that the data consumption on its network went past 20 GB/sub/month. T-Mobile USA also went past 10 GB/sub/month as the fastest growing network in the US. Overall data traffic is expected to more than quadruple over the next five years (Figure 4).

Furthermore, with automobiles coming onto the network, the amount of traffic per vehicle is going to explode. Practically every vehicle coming out of the factories will have an embedded cellular modem. The data demand for vehicle operation, entertainment, and communications in the vehicle is going to go up dramatically (Figure 4). All this requires effective management of the data traffic flows especially video which will continue its domination into the 5G era. Edge will play a significant role in processing data at the edge for not only offloading but also extracting intelligence and meaning out of data.

The confluence of these trends has meant that compute and storage doesn't have to be in a centralized server but could be highly distributed and the intelligence at the edge or in the device or the cloud could programmatically handle the load depending on the requirements. Additionally, the advent of 5G enables us to rethink the possibilities as the latency which has been a thorny problem for engineers for decades is getting a much-needed upgrade. The flat network architecture allows for sub 10ms latencies which makes addressing new class applications possible. The proximity of edge cloud

[2] Sources: Ericsson, GSMA

delivers a responsive and scalable solution. This architecture also becomes highly resilient during denial-of-service attacks or centralized cloud failures as each sub-segment of the Edge network can operate autonomously until the issues are resolved.

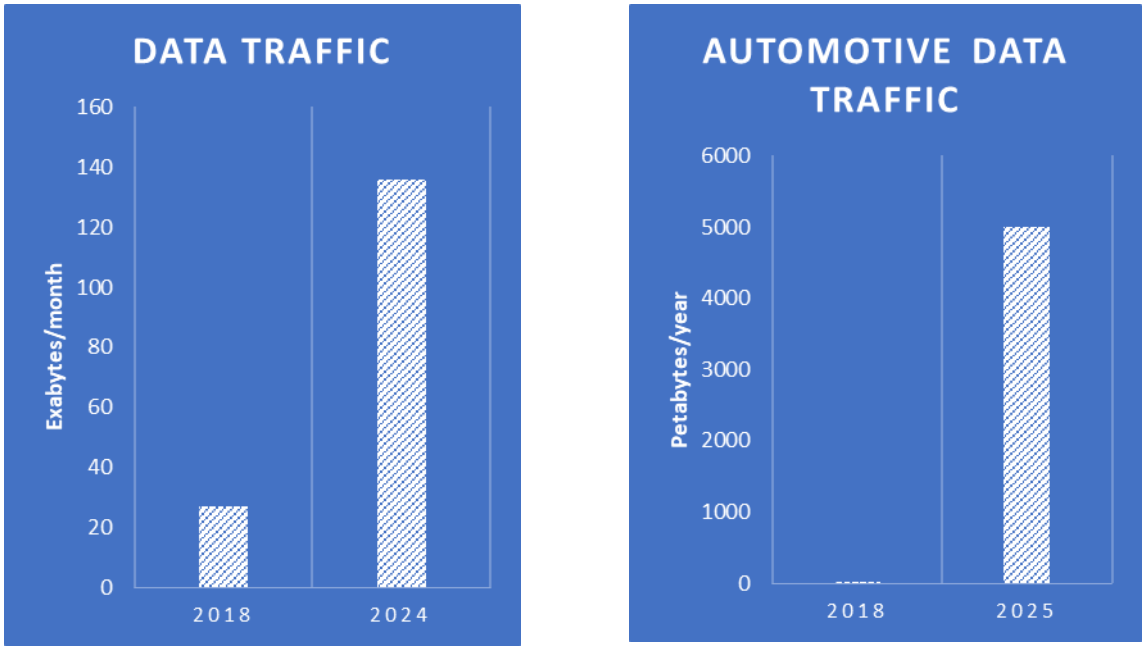

Figure 4. Growth in Mobile Data Traffic and Expected Automotive Data Traffic[3]

So, Edge Computing helps in redefining how and where compute takes place. We can instantly distribute highly computational tasks across multiple end-points and deliver intelligence and data at an order of magnitude faster than before. This makes VR streaming possible. One can now down thousands of facial recognition tasks at the edge in a matter of ms without uploading images to the cloud. Multiplayer gaming on powerful device can be accomplished because we know how to manage latency. We will be able to manage a complex mesh of autonomous vehicles with the help of onboard edge computing as well as interactions with the V2X infrastructure.

[3] Sources: Ericsson, AECC, Chetan Sharma Consulting

Existing Market Insights

The early days of Edge Computing will influence how developers and enterprises adapt to new ways of designing applications and services. There are several use cases we have discussed throughout the paper, but we will highlight a select few in more detail to give you a sense of the performance gains from Edge Computing and why it matters. In this section will highlight three important areas of imaging, website content delivery, and autonomous operations.

Imaging

It is accepted wisdom that responsiveness of applications and services has a direct correlation to improvement in consumer behavior, increase in commerce, and a step-change in revenues. As history of the tech industry of the last twenty years indicates, even slight upgrades can have a significant impact on user engagement. Edge computing will play a central role in providing significant uplift in improving end-to-end latency for not only current set of applications and services but more importantly the upcoming new breed of services and applications that will define them such as Virtual Reality, Multiplayer gaming, Facial Recognition, Enhanced Video Streaming, Automotive, Robotics, etc.

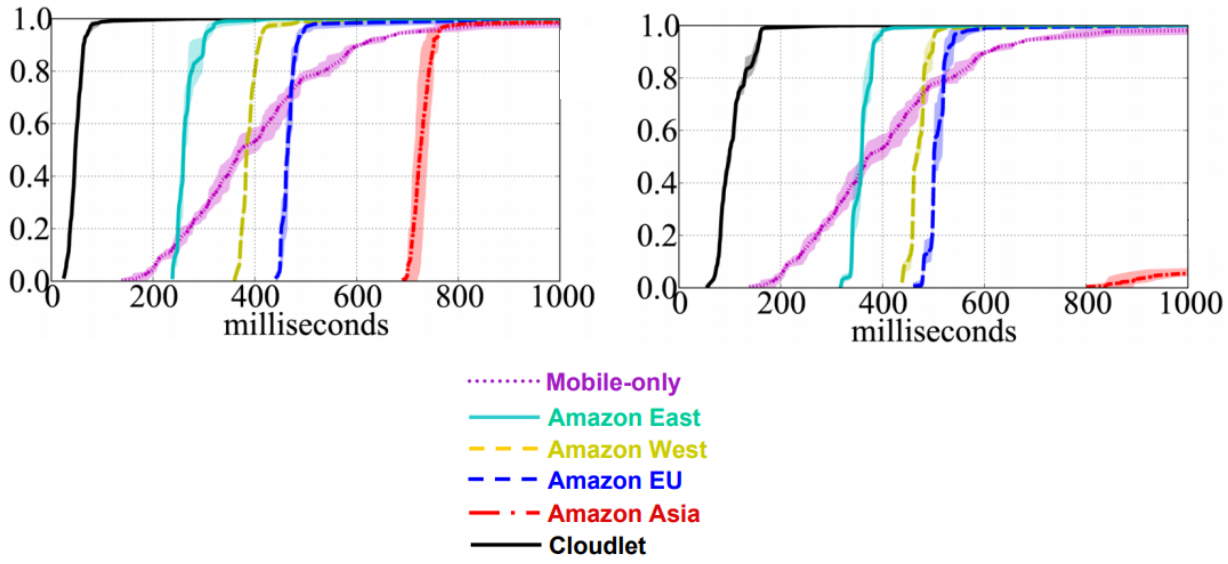

Figure 5. Response time distribution of an Augmented Reality application comparing Edge Computing vs. Cloud. The figure on the left is response times over WiFi while the one on the right is over the 4G LTE network[4].

[4] Edge Computing, A New Disruptive Force, Mahadev Satyanarayanan, SEC Keynote, 2017

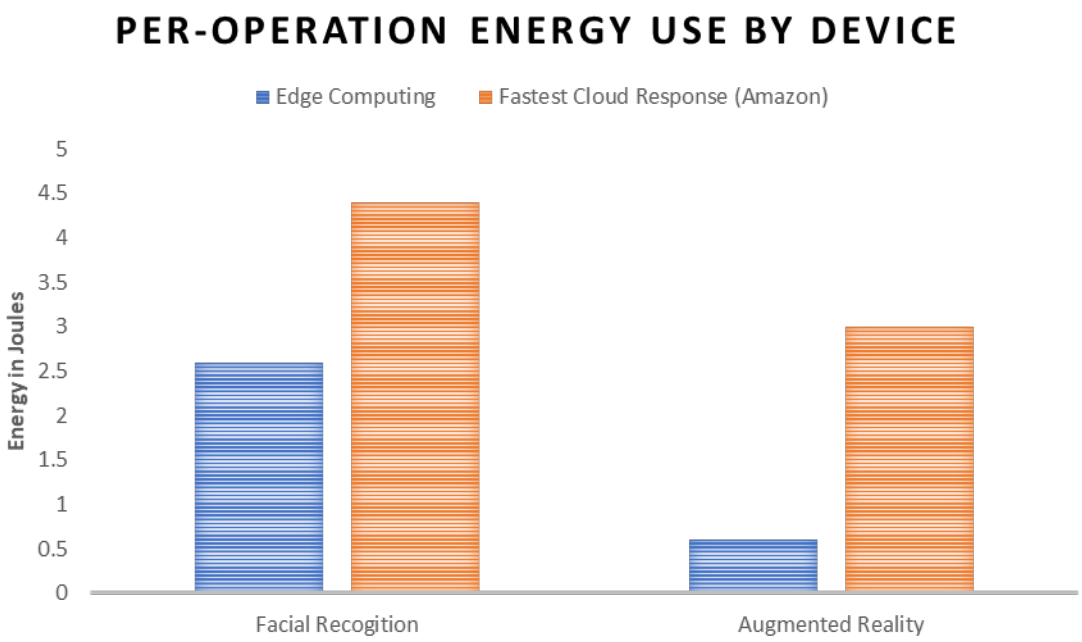

Figure 6. Per-Operation energy cost for Facial Recognition and AR applications on mobile device over WiFi to an Edge Computing Node vs. Amazon Web Services cloud.[5]

Measurement of response times for AR, Facial Recognition, and Video across cloudlets (Edge Computing cloud) vs. traditional cloud (Amazon Web Services) indicates there is a significant advantage of using near-edge cloud services compared to the cloud that is geographically dispersed. Figures 5 plots cumulative distribution function (CDF) of AR applications. Data clearly shows a marked improvement in response time using a cloudlet approach. Edge is not only great for response times but also battery operation on the devices as indicated in Figure 6 across two applications. Facial recognition is an application that performs well if the comparison of the facial vectors is done at the edge vs. the cloud. Similarly, the Augmented Reality applications and services need lower latency for optimal user experience. If the video frames aren't formed in time, the user can get dizzy and it will leave a bad taste.

We think applications that require a surge of computing resources and low latency will become natural first use cases for the industry. Once you add AI requirements, the traditional cloud architecture can't reliably serve the needs of the developers.

[5] Edge Computing, A New Disruptive Force, Mahadev Satyanarayanan, SEC Keynote, 2017

Content Delivery

Web has become the central pillar of commerce. The revenue created can be directly correlated to the response times of websites and applications especially during busy holiday season and one day commerce events such as Black Friday and Cyber Monday. The science behind website performance and conversions, abandonment rates, consumer sales, repeat traffic, and eventually revenue is well studied over the last decade. Enterprises have deployed different techniques to speed up the response times, be it the site design, data centers, or CDNs (content delivery networks). 74% of the mobile users will abandon a site after waiting 5 seconds for a page to load.[6] Each second shaved has a direct impact on user experience and the life time value of that customer. In fact, we would argue that the conversation in the Connected Intelligence Era has moved from response in seconds to content delivery in milliseconds, esp. for the millennials. The size of an average webpage has increased 10x over the last eight years so it is necessary to develop new techniques to improve responsiveness.

Mobile devices are increasingly playing a dominating role in commerce. Smartphones are driving almost 50% of the visits and almost one-third of the purchases.[7] As the number of objects on a webpage or an app increases, the demand for higher performance delivery will go up as well.

Edge Computing is powering a new way of distributing and serving the content to dramatically improve website performance. Startups like edgemesh are building a web acceleration platform that stores content at the very edges of the network to deliver performance not seen from traditional architectures. The distributed edge framwork of nodes and supernodes allows for edgemesh to distribute elements that form a webpage response closer to the edge device.

[6] https://www.akamai.com/us/en/multimedia/documents/infographic/web-performance-is-user-experience-infographic.pdf
[7] http://adobeenterprise.lookbookhq.com/adi2018

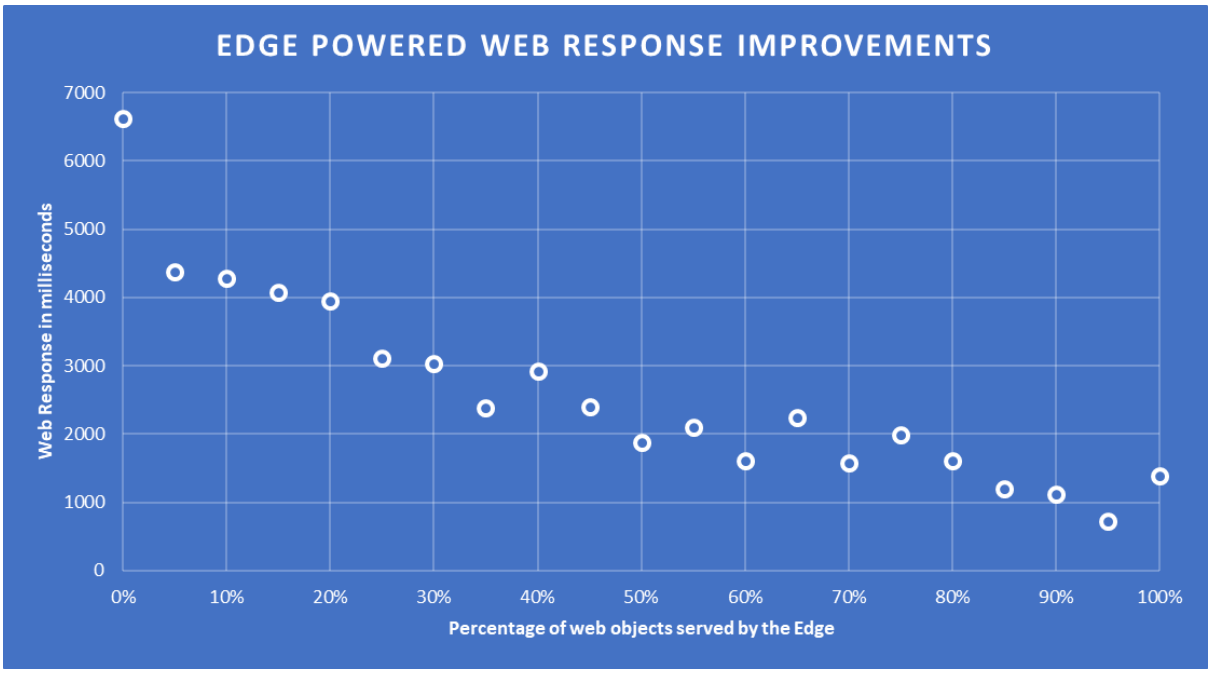

Figure 7. Correlation of Percentage of Web Objects served by Edge to Performance[8]

The analysis of field data shows why an edge architecture will become pervasive and a critical tool for developers worldwide. Figure x shows that as the number of objects on the page are served from the edge, the web response accelerates. In fact, the web response without any Edge support is roughly 6676 ms on average while the response with full Edge delivery is down to 1392 ms, an improvement that will have a direct impact on revenues.

Autonomous Platforms

The convergence of AI, 5G, IoT, Robotics, and Autonomous has meant that developers can reimagine industrial processes from the ground up. Functions that required humans can be automated because new toolset enables precision and performance that reconfigure the workflow be it robotic arms in a manufacturing plant, surveillance or maintenance drones inspecting an oil pipeline, or an autonomous fleet carrying goods from one corner of the country to another. In each of these scenarios, a highly reliable network is needed with very low latency such that the autonomous tasks could be synchronized, the information can be processed locally on the edge vs. the cloud so that intelligence gained from such analysis can be fed back into the loop to streamline the process.

[8] Data Source: edgemesh

As the number of edge devices increases, the complexity of delivering high fidelity network becomes important because any delay due to interference or packet collision needs to be mitigated in real-time. Furthermore, the gigabytes of data need to be processed at the edge so that meaningful insights can be delivered back to the operational workflow in an automated fashion in real-time. The efficiency gains can be 100-400% which has a direct impact on the bottom-line.

Energy companies are deploying drones for inspection and maintenance. Similarly, telecom and tower companies have the need to inspect their cellsites on a routine basis. In emergency situations, public safety is deploying drones to get real-time feed back to the command center. In all these cases, drones need to process enormous amounts of data, make sense of it while navigating their path to the desired destination. There will be occasions where the drone might not have network connectivity for certain portions of the flight plan. It will need a tight synchronization with the network to ensure that it is able to get network coverage maps in advance and is able to process and store data in the absence of an IP connection. Thus, Edge Computing becomes an integral part of the overall operation without which efficiency will suffer.

In a manufacturing plant that heavily relies on robotics and coordination of human tasks with robots on the floor, an extremely tight coordination is needed for not only better operations but also for safety compliance. Such environments require high throughput, reliability and extremely low latency. The more fluid the workflow, the higher the output and lower the error rate. Manufacturing industry is quite diverse from Chemicals to Motor vehicles, Plastics to Computer machinery, Textiles to Furniture. Each of these industry segments can be transformed by integrating elements of Edge Computing into the process.

It is a given that autonomous fleet will become the norm. Labor shortage and exploding demand makes the need for autonomous fleet more acute. We already see such vehicles operating in mining operations in Australia and airports in UK. Autonomous fleet requires synchronization of several key functions: real-time monitoring, end-to-end workflow management, maintenance and service, vehicular access, emergency response, and compliance and reporting. Each of these elements have different network needs from consistently high-throughput for information delivery and low latency of managing operations. Edge Computing plays a central role in management of data, intelligence, and the fleet.

How Edge Computing Will Impact Industry Verticals?

As is evident with the discussion above, Edge will play a central role in architecture and design for modern applications and services. While Edge Computing impacts both consumer and enterprise scenarios, the industry vertical can realize the gains much sooner. Essentially, Edge Computing allows the network, data, and intelligence resources to be available on-demand, as close to where they are needed as possible.

In this section, we will take a look at various industry segments and discuss some of the early use cases. Some of the industry segments are B2B2C so they indirectly impact the consumers for e.g. advertising and gaming where user experience drives consumer adoption, even a slight improvement in performance can yield significant gains in traction and revenue.

It should be noted that many of the use cases discussed below and throughout the paper are already happening. Entrepreneurs are not really waiting for a golden trigger point post which all the Edge Computing activities will commence but rather engineers are finding cases where Edge Computing can start making immediate impact like content delivery, gaming, agriculture, autonomous management, etc.

Below we briefly discuss how Edge Computing is shaping various industry verticals.

Table 1. Verticals and Edge Computing Scenarios

Verticals	Edge Computing Scenarios
Advertising	Advertising relies on accurate information about users and their context (location, etc.) to be effective in targeting. The integrity of the network and profile data is essential to have higher ROI in advertising. Additionally, there will be demand for more localized resource availability to do hyper-targeting which requires on-demand compute resources.
Gaming	Multi-player gaming requires tremendous coordination between devices that might be geographically dispersed. To have a good user experience, the user screens must move in sync. Even a few ms delay can turn off users and spoil the game. As such, accuracy of their location data and network latency becomes important.
Precision Agriculture	Precision Agriculture requires real-time input from atmospheric sensors and imagery from drones. This produces TBs of data that needs to be processed and analyzed to produce actionable recommendations on plans for irrigation, fertilizer, and disease management. By having this data analyzed in a matter of minutes instead of hours can help increase the crop yield by 7-10%.

Verticals	Edge Computing Scenarios
Manufacturing – Industrial Automation	Industrial automation will increasingly require robots working with humans which will need both highly-reliable and consistent throughput and low latency so the tasks can be effectively coordinated, keeping safety compliance in mind.
Connected Vehicles	As industry builds itself towards Level 5, the existing vehicles are being built with embedded connectivity and compute as well as aftermarket products that can provide similar functionality in the vehicle. These products can help in safe navigation, entertainment, and commerce.
Autonomous Vehicles	Consumer autonomous vehicles will be producing 40 TB/hour and this data will need to be digested and intelligence will need to be drawn from it. Sometimes, it will need to coordinate with nodes outside the vehicle that can operate on low latency. Communications and Compute will be central to the vehicle operation.
Autonomous Fleet	Autonomous Fleet within cities require coordination with other vehicular traffic as well as response from central facility that manages its movement. Platooned fleet requires close coordination of distance and highway speeds. As such availability of low-latency network nodes becomes vital.
Smart Cities	There are many smart city scenarios that can take advantage of Edge Computing but perhaps the most important one is that of security. Intelligent nodes can process data at the edge to detect problems before they happen and coordinate effectively with law enforcement when they do.
Oil and Gas	Oil rigs in ocean generate tremendous amounts of data but the network connectivity is expensive or incapable to deal with large amounts of data. It needs to be processed at the edge and the insights transferred when appropriate.
Wind Energy	Wind mills generate a lot of operational data about environmental inputs as well as how the individual units and parts are functioning. Usually, these wind farms are on ridges or remote areas. As such, data needs to be processed at the edge with exceptions send to a centralized node.
Retail	Retailers have varied communications and intelligence needs: understanding consumer behavior and movement in stores, analysis of who their customers are, where do they come from, which other stores they shop, etc. Having precise intelligence and contextual information is critical to forming engagement strategies.

Verticals	Edge Computing Scenarios
Stadiums – Customer Services	The sports franchises around the world pay special attention to season ticket holders who provide bulk of their revenues. Stadiums want to make their experience from start to finish a pleasant one – from finding a parking spot to getting them authenticated to go to their suite or seats to ordering meals. By having the context of customer at any given point, Stadiums can direct their customers to the right spot with minimum delay, use facial recognition to automatically check-them in and help them do commerce when in the stadium.
Entertainment Services	New technology enables new experiences. Instead of just watching the game from your seat or suite, if you could see the game from the eyes of players or refs on the field, it would elevate the experience tremendously. However, stitching dozens of images from live streams together in real time requires enormous computing and transmitting them in real-time requires high-throughput communications.
Disaster Relief	Disasters like earthquakes or hurricanes can be quite devastating to the network infrastructure. Computing and Networking equipment must be dropped in at a moment's notice. Given the scarcity of resources, it is more efficient to process information at the edge until full network connectivity is restored.
Healthcare	Edge Computing will have a transformative impact on health care from collecting and analyzing the real-time data from the devices on the patient to using VR during complex surgeries, education of new physicians, and transmission of large datasets.
Airlines	Each flight produces 1-5 TB of data. Data sometimes just sits there without getting processed. Transmission over cellular networks is cost prohibitive. Instead, the data should be analyzed in real-time and anomalies updated to ground crew and the cloud for further processing to reduce maintenance costs and delays.
Airports	Each day millions of travelers pass through thousands of airports around the world. There is a lot of "movement" data that is produced but hardly put to use. Customers can be directed to the right spot, airlines would have better idea about how many customers might miss their flights hence need help (allows them to resell their seats as well), shops can get more informed visitor data for better product placement and services, etc.
Drones	While industrial drones can operate without completely relying on the network, it needs

Verticals	Edge Computing Scenarios
	compute, AI algorithms for analysis, network coverage and availability intelligence data, and other conflict resolution data streams.
Smart Homes	The number of connected devices in a household is increasing every year and so is the throughput requirements in the house. Additionally, given the spread of sensors, requirements to manage security is growing. All this requires efficient collaboration between nodes, standards, and data streams.
Military	Military has to often setup a mesh network with each soldier or each node acting as a device. It must often transmit information including video streams at a very high data-rate and high fidelity. Many times, the information has to be locally analyzed in the mesh before it is transmitted to the cloud using slow satellite networks. As such bandwidth, latency, AI, and security become a necessary requirement for operations.

In this section, we discussed a number of use cases, some in more depth. The performance data clearly shows how Edge Computing environment can accelerate ROI for some of the application environments. In fact, applications like imaging, autonomous, and virtual reality won't even be possible without and Edge component aiding in delivery the bits to the end-devices. So, given the plethora of opportunities, how should we understand the marketplace that is multidimensional in nature? Where should developers invest their time and energy? Which services will be needed in the short-term (2-3 years) vs. long-term (3-6 years)? In the next section, we present a framework to evaluate Edge Computing opportunity landscape.

Introducing Edge Computing Framework

As we discussed before, the Edge Computing architecture allows us to bring the resources closer to where they are needed. This has direct impact on performance which results in better consumer experience and yield and also leads to the emergence of new use cases that were difficult to prove out with previous architectures. Different applications and services will need variety of resources at any given time. Figure 8 shows the overall Edge Computing framework. We consider a number of verticals and study them under two broad categories – Edge Services and Market Factors. Edge services provide us with a view of which Edge functions will be critical to the vertical and Market factors evaluate the opportunities for the vertical over the next decade.

Overall, we categorize the basic edge services as follows:

Compute – Providing a compute platform for data processing, content processing, and image processing. It also includes – management of video, offload of video, compression of data and video e.g. IoT, facial recognition, streaming services.

Access to Network Intelligence – Access to immutable network data and intelligence for key variables such as location, identity, traffic, and geospatial, e.g. multiplayer gaming, security, drones.

Latency – Dropping latency to a few ms can have a very positive impact on orchestrating the user-experience in several scenarios like multi-player gaming, automotive, and VR.

AI/Analytics – Availability of AI algorithms that can work independently or in concert with a distributed edge or distributed cloud architecture. Have the ability to complete the feedback loop in real-time e.g. advertising, autonomous vehicles.

Storage – Storage location for data and content e.g. multimedia content and enterprise content.

Data Residency - There are multiple scenarios whether driven by regulations (e.g. GDPR) or security (IoT Firewalls) or functionality (e.g. in agriculture) where the data can't travel too far from the source that produces it.

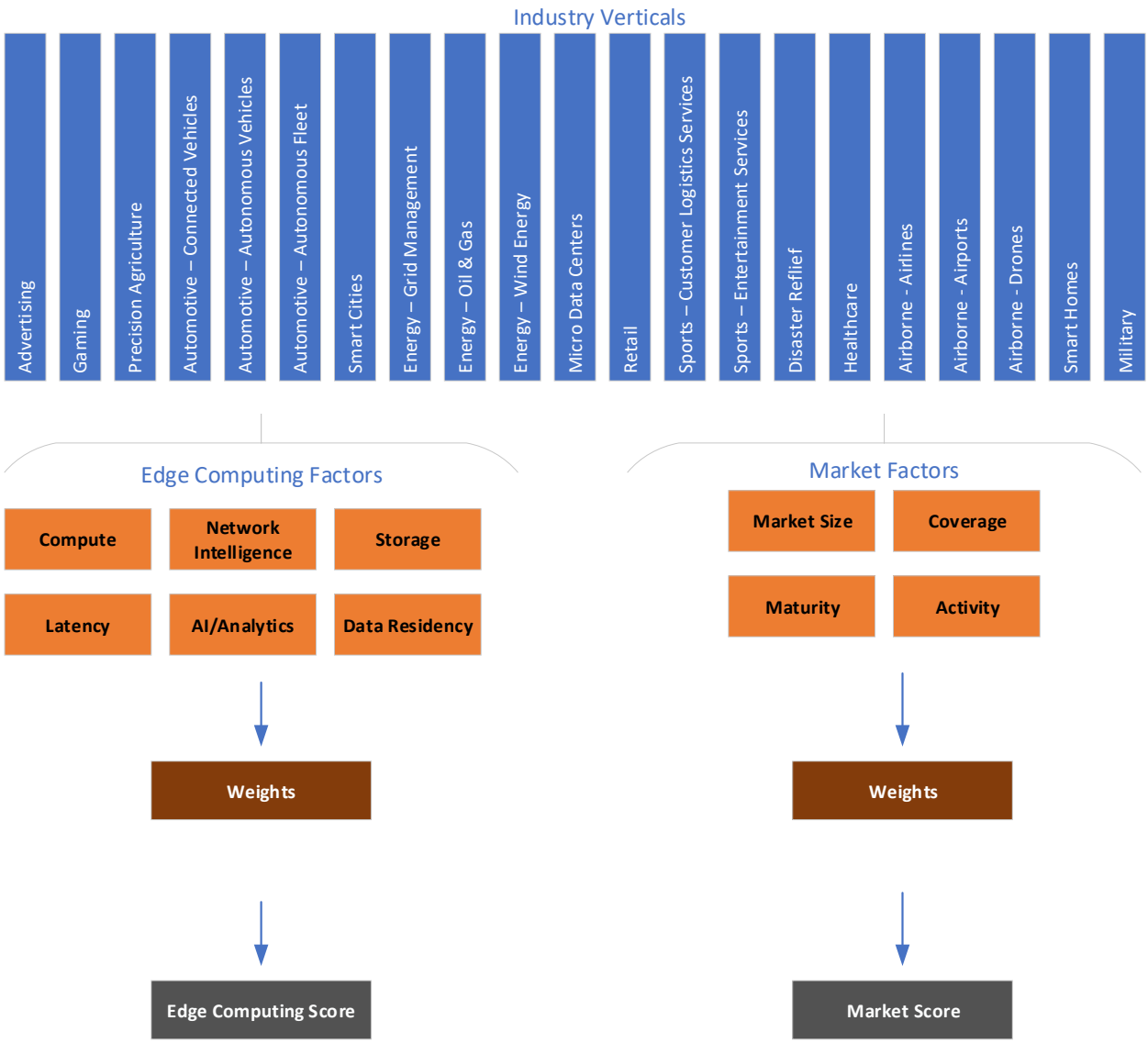

Figure 8. Edge Computing Framework[9]

Additionally, we need to consider market factors in ascertaining which application areas are likely to mature first based on industry's ability to absorb new technology and functionality, the coverage required to enable wide adoption, and the overall attractiveness of the opportunity.

Market Size – We already know that some areas such as video surveillance will be multi-trillion-dollar opportunity and are happening now but there are others on the horizon like Smart Cities and Connected Drones that could be used across multiple vertical industries

[9] Source: Chetan Sharma Consulting, 2019

Coverage required – Some applications are very localized like a factory or an oil pipeline and don't require nationwide coverage while others like multi-player gaming is best launched on a broad network across operators.

Maturity – Some industries by nature are conservative and are followers to the ones who are willing to invest and take the risk early on. For e.g. technology is adopted slowly in the healthcare industry while the security and financial industries are first ones to try out something that will give them an edge.

Activity – There are some segments that are receiving early traction and are indicative of potential success (doesn't always work this way but helps to have a pulse on where money is being invested and new startups are forming. This variable is also impacted by the geographies where the traction is taking place for e.g. getting a foothold in the US market is worth more to a developer than one of the emerging markets in Asia or Africa purely based on the potential monetary value. As such one must understand if the opportunity is regional or global in nature.

The variables described above presents one view of how the Edge Computing Opportunity Framework can be developed. One could argue that AI is a subcategory of Compute and similarly Data Residency is a sub category of Storage. Depending on the industry and the applications, developers will have their subjective view of how to analyze the Edge Computing market and develop their own perspective around the most important opportunities. We believe that AI, Analytics, and Data are critical elements of the Compute platform which can lead to new sub-ecosystems for e.g. one might gather weather data for agriculture but this micro-agriculture data could be used by other entities like City, Machinery Vendors, Infrastructure providers to develop their own models and insights from this dataset so the data can create its own economic zone with new business models. Furthermore, one can apply AI to gain insights and again mix with other information such as travel and transportation to create yet another set of business models and insights.

Similarly, on the market factors, necessity might be weighted heavily vs. the market size by certain industries like Military or Manufacturing. Whatever the permutations and combinations of the above discussed factors, using this framework allows us to map the opportunity with market demand.

Applying Weights

Not all factors are created equal. Some matter more than others so appropriate weights (Table 2) need to be applied to figure out the overall scores. Weights by definition are subjective. In this paper, we are providing our view of the world as it exists in early 2019 which will clearly evolve as more data becomes available from the marketplace. Each developer will have a different vantage point from which they can assess the Edge Computing ecosystem.

Table 2. Edge Computing Framework Variables

	Weight	Ranking Factor
Functionality (EC)-Centric		
Compute	0.20	1-5, 5 requires most compute
Access to Network Intelligence	0.25	1-5, 5 requires most NI
Latency	0.25	1-5, 5 requires lowest value of latency
AI/Analytics	0.20	1-5, 5 requires high-degree AI
Storage	0.05	1-5, 5 being most storage
Data Residency	0.05	1-5, 5 when DR is most important
Market-Centric		
Market Size	0.40	1-5, 5 being big market size
Coverage Required	0.25	1-5, 1 requires national coverage, 5 least coverage
Maturity	0.25	1-5, 5 being most mature
Activity	0.10	1-5, 5 being most active

Edge Computing Opportunities cut across a range of horizontal sectors and vertical industries. Some opportunities are very industry specific while others are across a range of verticals. For example, data analytics and AI are universal horizontal areas that will impact all vertical industries in some shape or form. Sometimes, AI is a subcategory under a horizontal area like Security, Robotics, Surveillance as it helps in analyzing and responding to data. AI is also a subcategory for almost all the vertical industries wherein data collected at the edge and otherwise is analyzed, processed, and fed back into the feedback loop to improve workflow, intelligence, and output. Table 3. gives a high-level view of synergies and the strength of relationship and dependencies of various horizontal elements and vertical industries.

One of the key areas of development is that of orchestration of workloads across multiple edge and centralized clouds. Even the intelligence and control points have to be distributed so as to maximize the efficiency of using the Edge Computing architecture. It will be become increasingly important the edge-enablement of applications and services is as simple as calling an API and the platform or the orchestration provides the brains to manage the requests, tasks, and responses. For applications such as multiple-user gaming where users are on different operator networks, it becomes critical that all operators in a given country work together to provide the abstraction layer to the developer, so the app has to call just one API in all instances and the orchestration layer, or the platform figures out the rest.

Table 3. Relationship strength of various industry verticals to horizontal capabilities/functionality[10]

	Data Analytics/AI	IoT	Video Surveillance	Video Streaming	CDN	Facial Recognition	Offload	Fraud Detection	VR Streaming	Security	Robotics
Advertising	●	○	○	●	○	●	○	○	○	○	○
Gaming	●	○	○	●	●	○		●	●	○	○
Precision Agriculture	●	○	○	○	○	○	●	○	○	○	○
Manufacturing – Industrial Automation	●	●	○	○	○	○	○	○	○	●	●
Connected Vehicles	●	○	○	●	●	○	●	○	○	●	○
Autonomous Vehicles	●	○	●	●	●	○	●	○	○	●	○
Autonomous Fleet	●	○	●	●	●	○	○	○	○	●	○
Smart Cities	●	●	●	●	○	○	○	○	○	●	○
Energy Grid Management	○	●	●	○	○	○	○	○	○	●	○
Oil and Gas	○	●	●	○	○	○	○	○	○	●	○
Wind Energy	○	○	●	○	○	○	○	○	○	●	○
Micro-data centers	●	○	●	○	●	○	○	○	○	●	○
Retail	●	●	●	○	○	●	○	●	○	●	●
Stadiums – Customer Services	●	○	●	●	○	●	○	●	○	●	●
Entertainment Services	●	○	○	●	●	●	●	○	○	○	○
Disaster Relief	○	○	●	●	○	○	●	○	○	●	●
Healthcare	●	●	○	●	○	●	○	○	○	●	●
Airlines	●		○	○	○	○	○	○	○	●	○
Airports	●	●	●	●	○	●	●	●	○	●	○
Drones	●	○	●	●	○	○	○	○	○	●	○
Smart Homes	●	○	○	●	○	●	○	●	○	●	○
Military	●	●	●	●	○	●	●	●	○	●	●

● – Strong Linkage ○ – Medium Linkage ○ - Weak or Emerging Linkage

Once we have understood the relationships, the weights of the framework, we are ready to study the output and how it maps the verticals on the opportunity map.

Analysis

Expanding the Edge Computing Opportunity Framework across the variables we discussed earlier in this section, we can calculate the Market Score and the Edge Score.

The Market Score indicates the opportunities from the market point of view that consider things like market size, traction within the VC, developer, and the ecosystem

[10] Source: Chetan Sharma Consulting, 2019

community, maturity of the industry vertical to be able to absorb and utilize Edge Computing, and the network coverage required to make it work.

The Edge Computing Score evaluates the various industry requirement needs against each variable that looks at the technical elements of compute, latency, data/AI, network intelligence, and storage.

After individually scoring for each variable, we apply the weights given to each variable to arrive at the final score for both the Market and the Edge. The output is shown in figure 9. One way to read the chart is to look at the relative attractiveness of a given vertical with other industries. Another way is to study the clustering of the industries that are Edge ready and have high market potential due to potential market size and ecosystem activity. These industries are ripe for Edge Computing use cases in the near to medium-term (upper-right quadrant).

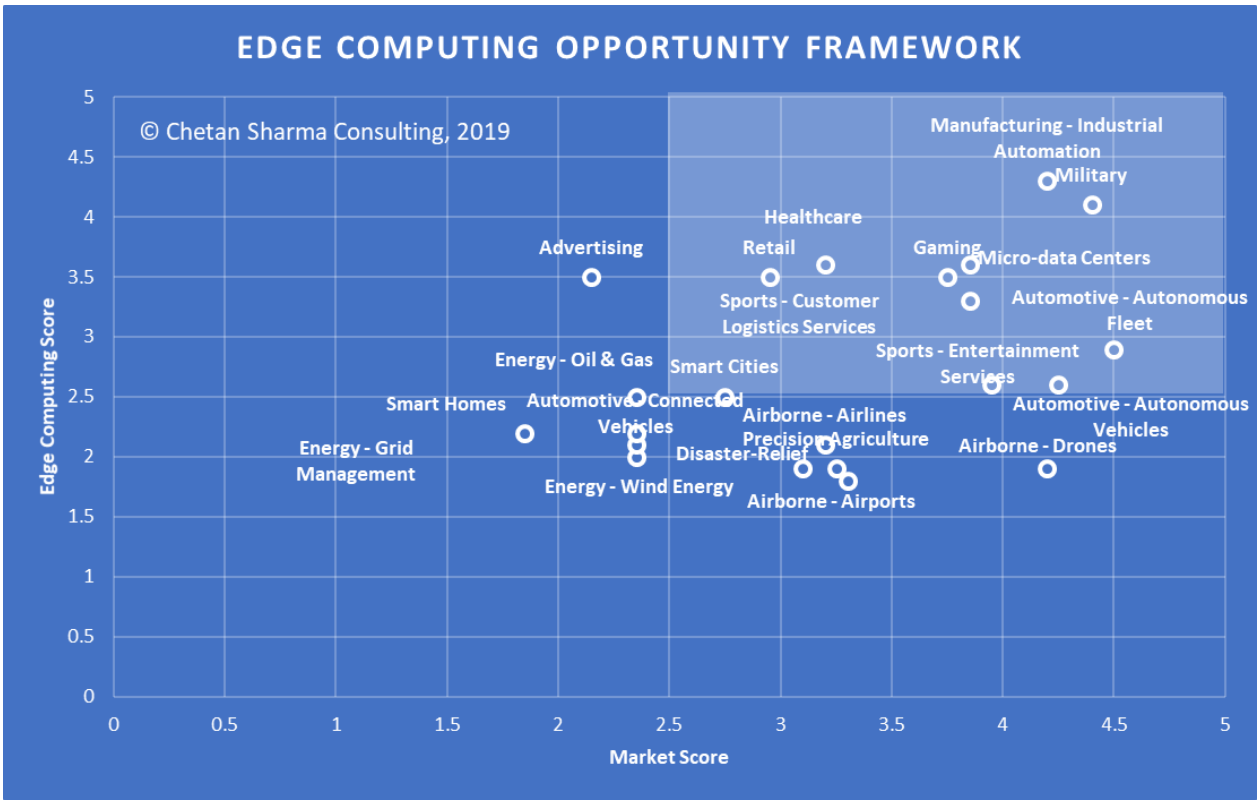

Figure 9. Edge Computing Opportunity Framework[11]

As we mentioned earlier, this is a subjective assessment which will both change with time as well as will be different for developers looking at the opportunities from a different lens. We hope that by providing an initial evaluation framework for Edge Computing, we can start an industry dialog and improve our collective understanding of

[11] Source: Chetan Sharma Consulting, 2019

the opportunity to collaborate, develop, and apply our learnings to build better Edge Computing use cases and implementations.

Given the distributed nature of Edge Computing, there will be multiple players who will be needed to make the fabric of this new computing paradigm work. We are clearly well into the journey of understanding how Edge Computing is going to shape the trillion-dollar industries and global GDP as a result.

Conclusions and Next Steps

Paradigm changes often happen gradually, and then, suddenly. We are during such change wherein the confluence of multitude of trends is enabling a rethink of network and application architecture. Edge will become the fabric of services just like Internet and mobile broadband did in the past and AI and 5G will become in the future. The numerous examples we discussed in the paper are testament to the fact that these changes are taking place now. One doesn't have to wait till every nuance has been ironed out but rather it is time to start reimagining the applications services to maximize the ROI.

To better appreciate the market and the opportunity, we have put forth a model framework of Edge Computing that can help us in prioritizing efforts, investments, and our collective understanding of the market. The paper presents a framework to derive at a set of answers. You will have a different view of the market depending on how your industry sees the opportunities. We invite you to play with the model and discover your own truths about the Edge Computing space. The model can be accessed at www.mobiledgex.com/navigator.

We look forward to your feedback and iterating on the model to learn from each other.

Acknowledgements

This research paper was supported by MobiledgeX. More information can be found at www.mobiledgex.com

About Chetan Sharma Consulting

Chetan Sharma Consulting is one of the most respected management consulting and strategic advisory firms in the mobile industry. We are focused on evolving trends, emerging challenges and opportunities, new business models and technology advances that will take our mobile communications industry to the next level. Our expertise is in developing innovation-driven product and IP strategy. Our clients range from small startups with disruptive ideas to multinational conglomerates looking for an edge. We help major brands formulate winning, profitable, and sustainable strategies.

Please visit us at www.chetansharma.com.

About the Author

Chetan Sharma is CEO of Chetan Sharma Consulting and is one of the leading strategists in the mobile industry. Executives from wireless companies around the world seek his accurate predictions, independent insights, and actionable recommendations. He has served as an advisor to senior executive management of several Fortune 100 companies in the wireless space and is probably the only industry strategist who has advised each of the top 9 global mobile data operators. Chetan serves on the advisory boards of Ericsson, Telefonica, Kymeta, NextNav, Zeotap, Opanga Networks, Mast and a number of other startups. Some of his clients include NTT DoCoMo, Disney, KTF, China Mobile, Toyota, Comcast, Motorola, FedEx, Sony, Samsung, Alcatel Lucent, KDDI, Virgin Mobile, Sprint Nextel, Skype, AT&T Wireless, Reuters, Juniper, Qualcomm, Microsoft, Google, Facebook, Mozilla, SAP, Merrill Lynch, American Express, and Hewlett-Packard.

Chetan is the author or co-author of a dozen best-selling books on wireless including *Mobile Advertising: Supercharge your brand in the exploding wireless market* and *Wireless Broadband: Conflict and Convergence*. He is also the editor of the *Mobile Future Forward Book* Series. His books have been adopted in several corporate training programs and university courses at NYU, Stanford, and Tokyo University. His research work is widely quoted in the industry. Chetan is interviewed frequently by leading international media publications such as *Time* magazine, *New York Times, Wall Street Journal, Business Week, Japan Media Review, Mobile Communications International,* and *TechCrunch,* and has appeared on NPR, WBBN, and CNBC as a wireless data technology expert. He is also the chief curator of the mobile thought leadership executive forums – Mobile Future Forward and Mobile Breakfast Series.

Chetan is an advisor to CEOs and CTOs of some of the leading wireless technology companies on product strategy and Intellectual Property (IP) development and serves on the advisory boards of several companies. He is also a sought-after IP strategist and expert witness in the wireless industry and has worked on and testified in some of the most landmark cases in the industry such as Qualcomm vs. Broadcom, Samsung vs. Ericsson, Sprint vs. Verizon, Openwave vs. 724 Solutions, and Upaid vs. Satyam. Chetan is a senior member of IEEE, IEEE Communications Society, and IEEE Computers Society. He has Master of Science degree in Electrical Engineering from Kansas State University and Bachelor of Science degree from the Indian Institute of Technology, Roorkee.