# Making the Edge Efficient, Scalable, and Sustainable

SUBZERO / WHITE PAPER

By Gordon Johnson

# INTRODUCTION

Each day, technology touches nearly every aspect of our lives in one way or the other. For example, how many times a day do each of us access one or more apps on our smart phone? This trend of needing, creating, transferring, and accessing data in fractions of a second isn't going away either. According to Gartner Research, by 2020, internet capable devices worldwide reached over 20 billion, and this number is expected to double by 2025. It is also estimated that approximately 463 exabytes of data (1 exabyte is equivalent to 1 billion gigabytes) will be generated each day by people as of 2025, that's the equivalent of 212,765,957 DVDs per day![1] Along with this increase comes the need to have this data as fast as possible, with minimum delay or latency, something most of today's data centers are not capable of.

The increase in data and the need for high-speed data transfers has inspired the recent trend known as edge computing. What exactly is the edge? What is an edge data center? How are edge data centers evolving and how can facility and data center managers be ready without being left behind? What about the challenge of making a resilient, modular, and scalable edge data center while maintaining high efficiency and reliability? This paper will answer these and many more questions about the edge in the following topics:

- What is an Edge Data Center
- The Evolution of Edge Computing
- How Organizations are Responding to Edge Data Centers
- Solving the Challenge of Modular and Scalable Edge Infrastructures
- Reliability and Efficiency Needed at the Edge
- Containment's Critical Role in Edge Deployments
- Bridging the Gap to the Edge, Now and Future

## WHAT IS AN EDGE DATA CENTER

Getting an exact definition of an edge data center may depend on whom you ask, but edge data centers are typically a smaller type or kind of data center than the core data center they support. They are also located close to the populations they serve. They typically connect to a larger central data center or even multiple data centers. But one thing all edge data centers have in common, they process data and services as close to the end user as possible, which allows the owner or the edge data center the ability to reduce data latency and improve the user experience.

While size can be one element in defining an edge data center, understanding the function and purpose of the edge helps us to better understand the meaning and need for edge data centers. Therefore, the main purpose of edge data centers, first and foremost, is to reduce latency and delays in transmitting data.

A good example is smart cars and autonomous driving, where decisions need to be made instantly and in real-time. What would happen if a car needed to make an emergency stop, but data from the car was transferred from one city to another (where a centralized data center is located) and back again. The results would be disastrous, because if the car needs to stop, you cannot wait that long for a response. This almost real-time speed of reaction can only be guaranteed if we have 5G and an edge data center in close proximity.[2]

Besides being located near areas they serve, edge data centers are also likely managed remotely. And while having many of the same components of a traditional data center, edge date centers are typically high-density computing environments packed into much smaller footprints with high kW/rack loads.

Size alone is not a defining factor for an edge data center. An edge data center is just one of many in a complex network that includes the main or centralized data center. The edge data center is responsible for housing mission critical data, applications, and services that need the data immediately. The edge data center simply moves the necessary computing close to where it's needed to accomplish this task.

## THE EVOLUTION OF EDGE COMPUTING


Why do we need edge data centers? The demand for computing closer to the point of data consumption, initially driven by the Internet of Things (IoT which is most synonymous with products like "smart homes and smart appliances", smart thermostats, smart phones, and anything with an IP address) has accelerated the rate and evolution of edge computing.

Presently, edge computing is leveraged by autonomous driving vehicles, drones, content delivery streaming, video monitoring services, augmented reality and virtual reality gadgets, and some artificial intelligence applications. Each require real-time data inputs and generate data that has value for a very short time frame. Edge enables filtration, analysis, and relaying the data output back to end users more quickly. This means bypassing a series of switches, routers, base stations, and other touchpoints that could make it a lengthy process and cost more time and resources.[3]

Take Artificial Intelligence (AI) and machine learning as an example, both of which are already playing significant roles in our lives by making complex tasks easier and simpler.[4] Our phones are already equipped with AI devices such as camera systems, voice to text translation, and facial recognition. Without the edge, waiting several seconds or more would be commonplace for simple everyday tasks. Through edge data

centers, companies reduce the physical distance that data needs to travel to reach a data facility, thus resulting in reduced latency. But this speed also necessitates large amounts of data transfers.

Currently, about 10% of data is created and processed outside a centralized data center or cloud, but according to Gartner, by 2025 this figure will reach 75%. According to Santhosh Rao, senior research director at Gartner, "Organizations that have embarked on a digital business journey have realized that a more decentralized approach is required to address digital business infrastructure requirements. As the volume and velocity of data increases, so too does the inefficiency of streaming all this information to a cloud or data center for processing."[5]

There are major benefits to decentralizing computing power away from the traditional data center and moving it closer to the point where the data is generated. While computing at the edge may have started as a gradual development and is still evolving, this demand for computing closer to the point of consumption represents a major shift in what types of services data centers will need to provide.

## HOW ORGANIZATIONS ARE RESPONDING TO EDGE DATA CENTERS

More organizations want it, eventually most will need it. So how should organizations move forward and respond to this growth? How and when will cloud providers move their workloads to the edge?

First, expect to see different types of edge data centers. For example, they'll be the local edge, the regional edge (larger but still in Metropolitan areas), and finally larger hyperscale data centers also at the edge. As for cloud providers moving workloads to the edge, it's estimated that there will be considerable growth in this area in the next 3 years, including big cloud and telco projects with hundreds of edge data centers needed.

Along with this increase comes the need for reliability. Reliable power and cooling is a priority, especially since many edge data centers will not have maintenance on-site but instead will be supported remotely. When considering reliability, if one of these edge data centers does fail, the nearest edge data center (perhaps just a few miles away) will likely take on some of the tasks. This means we can optimize our best people and use them at the best locations, not at the edge.

Who is going to build and run all these edge data centers in the future as demand grows? Telco and mobile operators need a solution to this answer, otherwise their 5G business and operations will not succeed. Fortunately, there are companies that are already offering edge data center solutions, including prepacked options with racks, power, and cooling. All the customer has to do is add outside power. Many of these solutions will combine edge functionality with sustainability solutions, as well as letting customers build and define their own edge, as will be discussed later.

Change is on the horizon, considerable edge projects are happening with numbers growing, and no doubt a few years from now the market will standardize around the successful growth of edge data centers. How can we meet and solve the challenge, both now and in the future, to make sure the edge is both modular and scalable?

## SOLVING THE CHALLENGE OF MODULAR AND SCALABLE EDGE INFRASTRUCTURES

As mentioned previously, more organizations want the edge and soon most will need it. But where will this infrastructure come from? Edge data centers must be resilient designs since we cannot expect someone to be there at a moment's notice. The key is to plan the architecture of the infrastructure properly in the first place.


Organizations will want to choose edge data center providers that allow the customer to build and define the edge as they go. It should be modular and scalable, that way they can add or subtract to their infrastructure as needed. It should be a fast, customized solution that is highly adaptable. Organizations will not and should not have their provider define the edge for them, they'll want to define their own edge based on their specific needs and application.

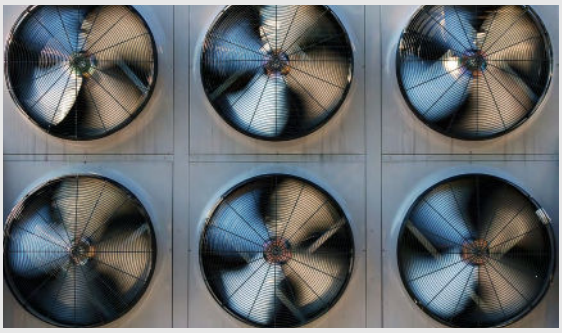
Besides adaptability, speed to market will also be essential for lowering customer costs. Edge data centers will need to be built quickly, delivered on-site, and be easily expandable. We're not referring to one of those 1-rack edge data centers, but to a completely designed and built-out infrastructure that can be drop shipped and immediately deployed in a matter of weeks.

And again, if expansion is needed, the edge infrastructure should be such that a second or third unit can quickly and easily be added, regardless of the location or how long it has been in operation. If one unit needs to be removed and relocated somewhere else, no problem. Being modular, scalable, and flexible should be a priority of any edge data center provider. When it comes to the edge, organizations should want and expect to have the edge their way and how they want it, not how some edge data center provider tells them it should be.

## RELIABILITY AND EFFICIENCY NEEDED AT THE EDGE

For large traditional data centers, the annual cost of electricity used in cooling can easily exceed the cost of electricity used by the IT equipment in processing. The ratio between the two is called PUE (Power Usage Effectiveness) and is a measure of the total energy consumption of a data center. Theoretically, it should cost businesses less to cool and condition several smaller edge data centers than one big traditional data center. In addition, due to the ways in which some electricity companies handle service area billing,

the cost per kilowatt may go down for the same server racks hosted in several small facilities rather than one big one.[6]



To deliver the most efficient edge environment as possible, we're putting data centers where they've never been, so we cannot rely on unlimited power and unlimited cooling. Reliable power and cooling are a must, especially since many of these data centers will not have personnel on-site due to remote locations and the sheer quantity of edge data centers. Reliability, efficiency, and sustainability should force us to rethink how we design and deploy the edge.

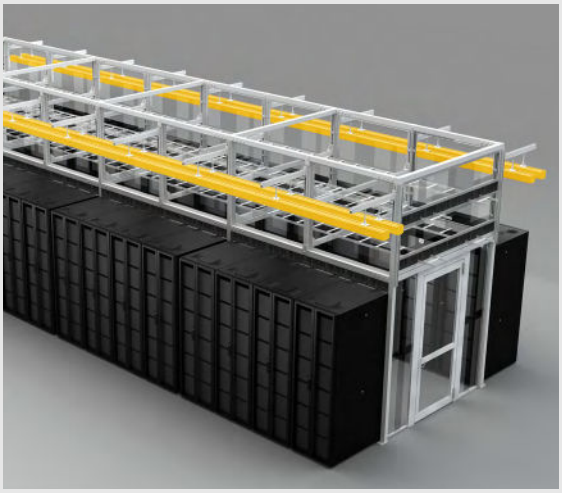
In traditional data centers there can be a lack of incentive for IT personnel to think or consider efficiency and what it means in terms of lowering energy costs and carbon footprint. With the edge, everyone should be thinking about efficiency challenges holistically, we need to consider the big picture! We should look for edge designs and providers that are invested in reducing the need for power and cooling, while maintaining IT equipment reliability. The impact on efficiency and sustainability depends on this.

Looking ahead there is an edge energy challenge coming. Energy consumption by edge data centers is predicted to exceed 3K TWh per year by 2040, which is currently the equivalent energy consumption of 275 million U.S. households. It's also estimated that by 2040, 80% of total data center energy consumption will be from edge data centers.

When we discuss data center efficiency and resource consumption, the focus is usually only on electricity which impacts total operating costs, but electricity also impacts the environment, so it's worth thinking about the impact of water consumption on data centers and their environmental impact as well.[7] What can help the edge be energy efficient, environmentally responsible, reliable, and sustainable all at the same time?

## CONTAINMENTS CRITICAL ROLE IN EDGE DEPLOYMENTS

The future demand for edge data centers and the ensuing power consumption means that it is our environmental obligation to look at sustainability.[2] The easiest way to do this is with containment. Containment has often been described as a "no brainer" decision when it comes to data centers. It's the easiest way to save money and increase efficiency in the data center. Containment also makes the data center an environmentally conscious place, because instead of consuming energy, containment saves energy. This is especially true with edge data centers.



Edge data centers will have smaller footprints but higher rack densities. Containment will help you get the most out of an edge deployment. Because containment prevents cold supply air from mixing with hot exhaust air, the supply temperatures at server inlets can be increased and kept at the desired level throughout the data center. Since today's servers are recommended to operate at temperatures as high as 80.6°F (27°C), containment allows for higher supply temperatures, less overall cooling, lowering fan speeds, increased use of free cooling, and reduced water consumption, all important factors for improving efficiency and lowering the carbon footprint.

The edge with containment is energy-conscious because it consumes less power than without containment. This, in turn, means less overhead since edge cooling requirements are reduced. The edge with containment operates without leaving a hefty impact on the environment. An environmentally friendly edge data center is always a cost-effective edge data center.

In addition to saving energy, containment at the edge means longer Mean Time Between Failures (MTBF) of the IT equipment, lower PUE, and more. The net effect of containment is a high efficiency edge data center, lower carbon footprint, and a green edge data center.

Since containment reduces the amount of air conditioning or cold supply air (CFM) needed to cool the IT equipment, it should be an integral part of any edge data center deployment. If you're not planning on containment in your edge data center, don't even

consider HPC or high-density racks. Without containment, you'll need to flood the room with an excess of cold supply air, and edge data center managers can expect lower kW per rack densities, potential hot spots, and higher energy bills.

## BRIDGING THE GAP TO THE EDGE, NOW AND FUTURE

Earlier it was discussed that edge data centers will need to be modular, adaptable, and built quickly. Additionally, edge providers shouldn't define the edge, instead the customer should be able to define their own edge and what they expect from it. So how do we get to this point, both now and in the future?

First, look for an edge deployment that provides a quick and easy installation. When some people think of an edge data center, they think of a highly engineered container solution that needs to be specified and assembled in a warehouse, then drop shipped to its location. Instead, some providers now offer an enclosure built on-site that can go up in a matter of days, with ground supported or ceiling hung infrastructure that supports ladder racks, cable trays, racks, cooling, etc. The customer chooses their own power and/or cooling vendor ahead of time, so all that's needed is for the customer to provide the IT stack and power everything up.

This type of scalable infrastructure includes conveyance, racks, and of course containment for maximum efficiency and reliability. Customers that will benefit from this type of edge deployment include colocation companies and hyperscale data centers that want and need the edge on-site, large scale e-commerce that might normally not have their own IT infrastructure, and companies that want their own small on-site data center to name a few.

Most edge data center providers try to standardize what the customer will get in terms of cooling, power, and rack density. Because of this, customers have a hard time finding what they want and are settling for what a provider is telling them they need. This should not be the case. The speed and ability for customers to find what meets their needs and requirements is what the edge needs to be. They need their edge deployment to be super flexible and in operation in a matter of weeks.

Bridging the gap to the edge needs to be easy, fast, flexible, energy efficient, and low-cost. Make sure any edge deployment checks off all these boxes and more.

## Conclusion

Today's data centers move high amounts of data, at times thousands of miles away from end users, but that's changing. Demand at the edge is increasing dramatically requiring more edge deployments, estimated at 12-15% increase year over year. To meet this demand, the Essential Micro Data Center (EMDC) from Subzero Engineering is the perfect solution.

The Subzero EMDC is a highly efficient, reliable, and scalable edge data center that the customer defines instead of the provider defining for them. It comes equipped with everything needed except the IT stack and outside power. It is also equipped with full containment to separate the cold supply from the hot exhaust air to maximize energy efficiency, reduce energy costs, and lower the carbon footprint.

The Subzero EMDC is a game changer when it comes to scalable edge infrastructures, both now and into the future.

# ABOUT THE AUTHOR

Gordon Johnson is the Senior CFD Engineer at Subzero Engineering, and is responsible for planning and managing all CFD related jobs in the U.S. and worldwide. He has over 25 years of experience in the data center industry which includes data center energy efficiency assessments, CFD modeling, and disaster recovery. He is a certified U.S. Department of Energy Data Center Energy Practitioner (DCEP), a certified Data Centre Design Professional (CDCDP), and holds a Bachelor of Science in Electrical Engineering from New Jersey Institute of Technology. Gordon also brings his knowledge and ability to teach the fundamentals of data center energy efficiency to numerous public speaking events annually.

# REFERENCES


1. How Much Data Is Generated Each Day?, Posted on 15 April, 2019 by Jeff Desjardins; Infographic: How Much Data is Generated Each Day? (visualcapitalist.com)
2. The Edge & Edge Data Centers: Gaining Clarity – Doteditorial, Posted on November, 2019 by Bela Waldhauser; The Edge & Edge Data Centers: Gaining Clarity - doteditorial - On the Edge: Building the Foundations for the Future - Issues - dotmagazine
3. Transitioning From Data Centers To Edge Data Centers: The Next Chapter, Posted on 6 January, 2021 by Preetipadma; Transitioning from Data Centers to Edge Data Center: The Next Chapter (analyticsinsight.net)
4. What Is Edge AI?, Posted 1 May, 2020 by James Gordon; What is Edge AI and How Will it Change Things? | GCU Blog | GCU Blog
5. What Edge Computing Means For Infrastructure And Operations Leaders, Posted on 3 October, 2018 by Rob van der Meulen; What Edge Computing Means for Infrastructure and Operations Leaders - Smarter With Gartner
6. ZDNET What Is Edge Computing? Here's Why The Edge Matters And Where It's Headed, Posted on 1 June, 2020 by Scott Fulton III; What is edge computing? Here's why the edge matters and where it's headed | ZDNet
7. Datacenter Energy Consumption, Posted on 19 December, 2020 by DatacenterEdge Nate; Datacenter Energy Consumption - Datacenter Edge - Industry news