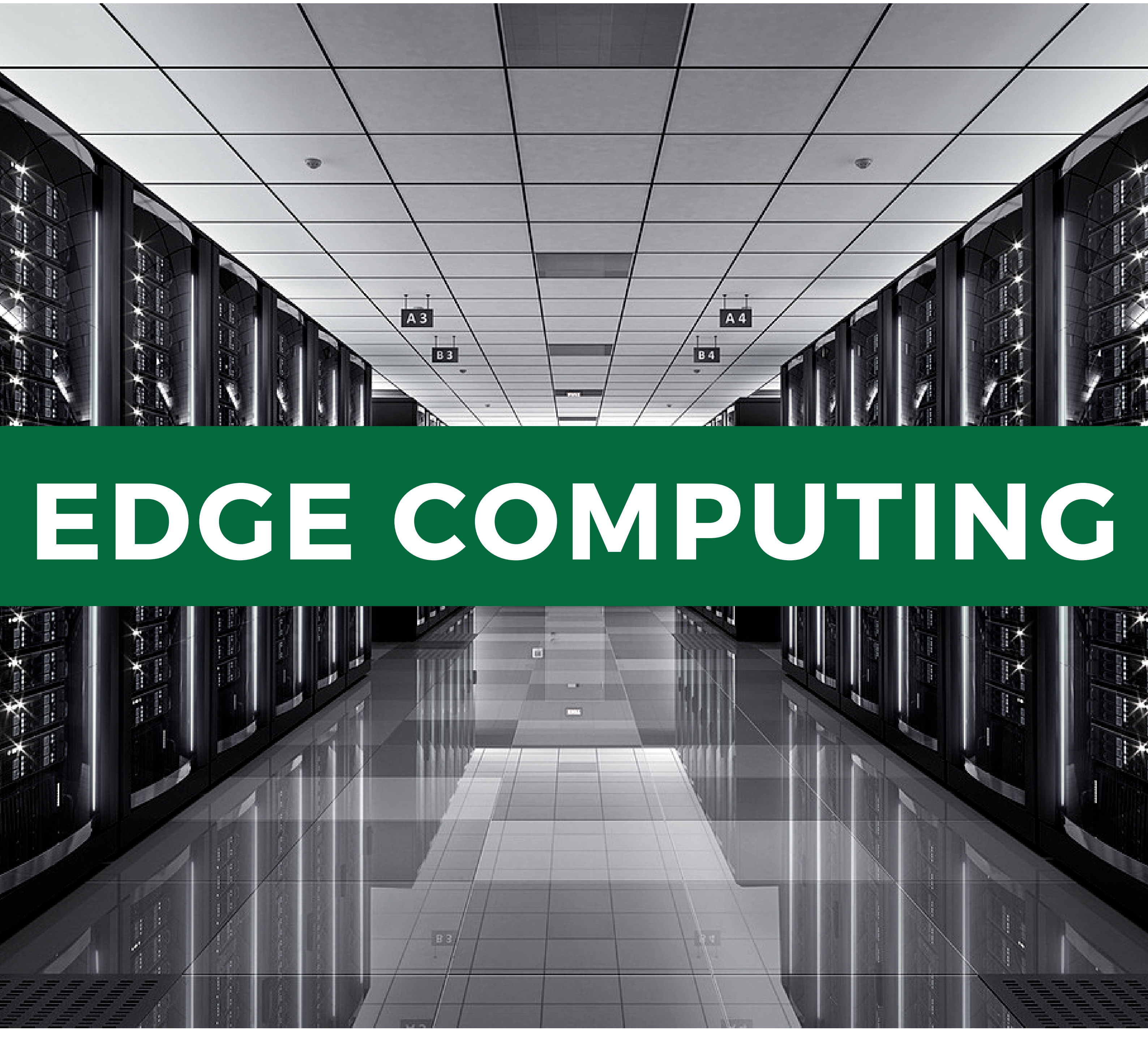
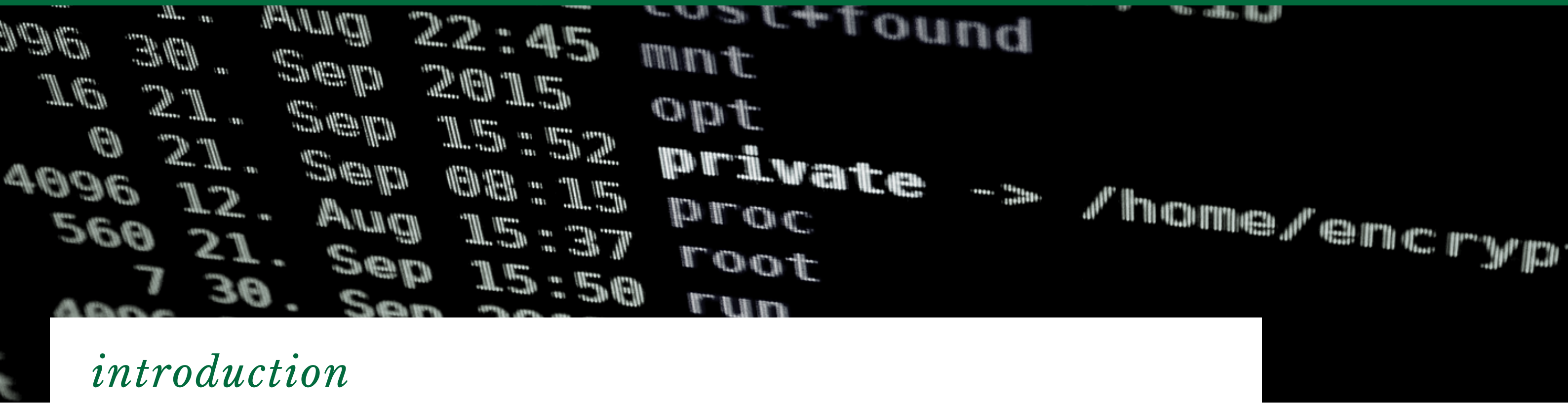


EDGE COMPUTING

HOW LEADERS SHOULD EVALUATE INVESTMENT

PREPARED BY
Alex Pandrangi 'T'19



introduction

WHAT IS EDGE COMPUTING?

With ever increasing literature around the adoption, strategic importance, and projected growth of cloud computing, we must also acknowledge that not every computing use-case makes sense for traditional cloud infrastructure. What happens when a slight rise in latency rates can destroy hundreds of millions of dollars in value by disrupting a high frequency trading algorithm? How would the public react if an autonomous vehicle crashed and killed a passenger because it had not yet retrieved the latest traffic data from the cloud. Would you feel safe if the military performed all battlefield operations through the cloud when often operating in areas with limited network connectivity?

For certain use-cases, traditional cloud infrastructure is not ideal. For many of these use-cases, edge computing is becoming a more viable and appropriate option.

EDGE COMPUTING IS THE PRACTICE OF PROCESSING DATA NEAR THE EDGE OF THE NETWORK, CLOSER TO WHERE THE DATA IS BEING GENERATED, INSTEAD OF IN A CENTRALIZED DATA-PROCESSING WAREHOUSE.

The edge of the network could refer to the actual device or sensor level, which would represent precisely where the data is collected. Or, it could also refer to very nearby the data collection, such as at a micro data center set up to process the information being generated. Regardless of implementation, edge computing essentially upends the paradigm of needing to transmit collected data long distances before any meaningful analysis can be performed or insights generated.

FIVE FACTORS FOR LEADERS TO CONSIDER

CRITICAL DATA VOLUME & VELOCITY OF ITS GROWTH	The higher the volume and velocity of growth, the greater investment consideration edge computing should receive.
TIME SENSITIVITY OF CRITICAL DATA FOR DECISION MAKING	The shorter the required feedback loop for incorporating critical data into the decision-making process, the greater investment consideration edge computing should receive.
NETWORK RELIABILITY WHERE DATA IS COLLECTED	The more rugged the environment and less reliable the network infrastructure, the greater investment consideration edge computing should receive.
COST OF IMPLEMENTATION	The greater the return on investment (ROI), the greater investment consideration edge computing should receive. Leaders should be aware that safety concerns can complicate ROI calculations for certain applications.
DATA SECURITY	Investing in edge computing has both benefits and drawbacks from a data security perspective – the net effect depends on the specific application and organization.

KEY INVESTMENT CONSIDERATIONS

Bringing computing power closer to the data source has critical advantages, and in a world of ever-increasing sources and velocity of data that can now be continuously collected, use-cases and applications for edge computing seem to be expanding. Edge computing thrives when any of the following dynamics are at play. Leaders must evaluate their organizations and/or intended applications along these three dimensions in order to determine whether an investment in edge capabilities is justified.

1 HIGH CRITICAL DATA VOLUME & VELOCITY OF ITS GROWTH:

At a certain point, existing and even expanding network infrastructure will not be able to transmit ever growing data loads. Through implementation of edge computing, a large percentage of overall network data traffic can be reduced through processing data near its collection and only relaying the critical outputs.

2 REAL-TIME ANALYSIS OF CRITICAL DATA FOR IMMEDIATE IMPLEMENTATION INTO DECISION MAKING:

Edge computing provides an opportunity to reduce latency substantially, allowing significantly more effective optimization of decision-making processes in real-time.

3 LIMITED OR LACKING NETWORK CONNECTIVITY:

When connectivity is limited or could become limited unexpectedly, edge computing allows for the continuation of data processing when there otherwise would be a noticeable interruption, which is simply not acceptable for certain use-cases.

HIGH CRITICAL DATA VOLUME & VELOCITY OF ITS GROWTH

As the number and sophistication of deployed devices grow, the number of data streams and the velocity with which that data is collected are also increasing rapidly. IDC estimates that total data generated worldwide will grow from an estimated 33 zettabytes in 2018 to an astounding 175 zettabytes by 2025.

At a certain point, existing and even expanding network infrastructure will not be able to transmit ever growing data loads. Further, there is cost associated with transmitting data across networks, which scales (to some extent) with the amount of data transmitted.

> WITH EDGE COMPUTING, THE DATA COLLECTED CAN BE PROCESSED AT THE EDGE OF THE NETWORK, ONLY RELAYING THE MOST CRITICAL DATA BACK TO A CENTRAL STORAGE REPOSITORY FOR HISTORICAL ANALYSIS OR CONTEXT.

Through implementation of edge computing, a large percentage of overall network data traffic can be reduced through processing data near its collection and only relaying the critical outputs.

REAL-TIME ANALYSIS OF DATA FOR IMMEDIATE DECISION MAKING

The use of data to improve business decision making has been a key tenet of the cultures of many successful digital-savvy firms. As discussed above, however, the total amount of data being produced over any given time frame is often overwhelming for a human to process and analyze. Further, there are many use-cases where decision making must be augmented in real-time with little room for error. In these instances,

> THE LATENCY IN TRANSMITTING DATA BACK TO A CENTRAL SERVER OR REPOSITORY, PERFORMING THE COMPUTATIONS THERE, AND THEN RELAYING INSTRUCTIONS BACK TO THE POINT OF IMPLEMENTATION CAN CONSTITUTE THE DIFFERENCE BETWEEN SUCCESS AND FAILURE OR PROFIT AND LOSS.

Edge computing provides an opportunity to reduce latency substantially, allowing significantly more effective optimization of decision-making processes in real-time.

LIMITED OR LACKING NETWORK CONNECTIVITY

There is also a more fundamental limitation that may require some form of edge computing: limited network connectivity. Telecom infrastructure is far from perfect and never will be, especially with many governments controlling, either directly or indirectly, the development of their nation's telecom infrastructure.

Additionally, certain harsh environments, such as deserts or disaster areas, do not lend themselves to maintaining traditional, reliable network infrastructure. When connectivity is limited or could become limited unexpectedly, edge computing allows for the continuation of data processing when there otherwise would be a noticeable interruption, which is simply not acceptable for certain use-cases.

WHERE IS EDGE COMPUTING APPROPRIATE?

Based on the three factors discussed above, the following spectrum highlights where selected use-cases stand on a relative basis between optimally suited for traditional cloud IT infrastructure versus edge IT infrastructure.

TRADITIONAL CLOUD/IT INFRASTRUCTURE → **EDGE IT INFRASTRUCTURE**

Higher Education Virtual Desktop/VPN

Website & Email Hosting

Critical Data Storage & Backup

Other non-time-sensitive business applications

Consumer Drones

Tractors & Agricultural Equipment

Military Drones

High Definition Video Streaming

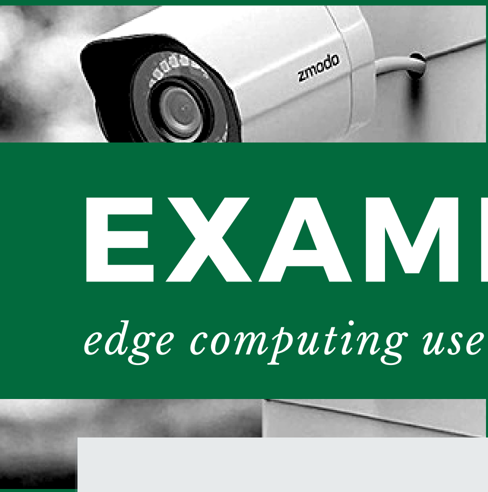
High Frequency Trading

Blood & Vaccine Monitoring

Autonomous Vehicles

Battlefield Analytics

(CENTRALIZED IT MODEL) → **(DISTRIBUTED IT MODEL)**



EXAMPLES

edge computing use-cases

SURVEILLANCE CAMERAS

Consider a collection of high definition cameras spread across a university campus. If these cameras are simply constantly recording and sending all of the video content to a central repository, then there is an incredible amount of data being sent over the network. However, if the devices themselves can perform some of the processing, then the amount of data being sent can be limited by a significant factor. For example, if the cameras have the processing ability to determine whether or not there is a human in the frame and subsequently record and send back to the central warehouse only video that contains humans, this accomplishes the goal of maintaining appropriate surveillance while substantially reducing the data sent across a network back to a central warehouse.

MANUFACTURING

In many industrial technology manufacturing processes, there are sensors continuously collecting data on overall efficiency, accuracy, equipment performance, etc., all in real time. For high-end products, every batch can be extremely valuable. Let's consider the production of sapphire glass, which is used in certain high-end watch displays, among other applications. As part of the production process, crystals are melted down, so they can eventually be reshaped. This melting occurs in a furnace, which must operate within a specific temperature band in order to create an acceptable product for the next step in the process. The temperature of the furnace is consistently monitored by an industrial thermometer. With edge computing capabilities at the system level, the temperature of the furnace can be seamlessly adjusted as any change is registered by the thermometer. With traditional cloud computing capabilities, the temperature information is sent back to a central repository for processing and then the required adjustment must be sent back and implemented within the system. By reducing this latency, edge compute power can decrease shut-downs and increase output, contributing to better utilization rates, decreased waste, and higher profits in smart manufacturing processes.

AUTONOMOUS VEHICLES

Autonomous vehicles contain an incredibly large number of sensors that generate data about the world around the vehicle. The data is processed, analyzed, and converted into actions. If this data had to be sent to the cloud for analysis and relayed back to the vehicle, it would create an extreme safety hazard. Any time the external network connection became limited, the vehicle would not function properly (not to mention the latency issues already discussed). For scenarios and applications involving limited network connectivity, some form of edge computing capability is a must rather than a measured improvement.

ADDITIONAL CONSIDERATIONS

4 COST OF IMPLEMENTATION:

There are many applications where investment in edge computing can be considered on a return on investment basis. The camera and smart manufacturing examples discussed above would fall into this category. Increasing compute power within edge devices represents an increase in cost for each device. This investment must be more than offset by some form of cash flow generated by the edge enablement, reduction in cost of network data transmission in the case of the camera example and increase in profit from greater sapphire glass yields in the manufacturing example.

For some applications, however, cost cannot serve as the only determining factor guiding investment decisions. These cases typically involve peoples' safety. Autonomous vehicles fall squarely in this category. If significant computing capabilities are not embedded within the autonomous vehicle itself, then it has no chance of being safe or, consequently, of being adopted broadly. For these types of applications, edge computing is a necessity, not an option. Cost and return on investment can only be a real consideration factor in determining whether to undertake the whole project itself in these instances.

5 DATA SECURITY:

Shifting from a more traditional cloud infrastructure toward edge computing presents some complexity around data security. It is certainly difficult to provide the same level of security at the edge of the network as can be focused on a more centralized location, as the same total resources dedicated to security must be allocated across many disparate edge locations in addition to the central repository. And, when retaining more meaningful data and processing power away from a central repository, this presents more points of entry for data-compromising attacks. These are very real drawbacks for edge computing from a cyber security perspective.

There is at least one major benefit to a more distributed model. While there are more entry points for would-be attackers to gain access to valuable data, presumably the data that could be accessed at any one edge entry point would represent a much narrower subset of data than if one large repository were somehow compromised.

From a net perspective, most organizations will likely view data security around a distributed model as more of a challenge than a benefit when compared to the status quo of traditional cloud infrastructure. However, some companies may view the distributed model with robust data partitioning to be more appealing. In either case, the overall benefits from a distributed model (discussed in previous sections), could plausibly exceed the security drawbacks. Because of the gray area, leaders should involve their cyber security team members early when considering investment in edge computing.

KEY QUESTIONS FOR LEADERS

INVESTMENT CONSIDERATION	QUESTIONS	ANSWERS THAT FAVOR INVESTMENT IN EDGE COMPUTING
CRITICAL DATA VOLUME & VELOCITY OF ITS GROWTH	• Which data is the most critical to managing the business? • How much critical data are you dealing with? • How fast are critical data streams growing?	• **Higher critical data volumes** • **Higher rates of growth of critical data**
TIME SENSITIVITY OF CRITICAL DATA FOR DECISION MAKING	• How time sensitive is critical data? • How is data processed? • What does the feedback loop look like in terms of learning from and implementing changes based on data?	• **Greater time sensitivity of data** • **Shorter required feedback loop for acting on critical data**
NETWORK RELIABILITY WHERE DATA IS COLLECTED	• In what environment is the data being collected? • How reliable is the network infrastructure?	• **More Rugged Data Collection Environment** • **Less Reliable Network Infrastructure**
COST OF IMPLEMENTATION	• Can application be evaluated on incremental ROI basis compared to traditional infrastructure? • Or, is the application either edge-based or not at all?	• **Higher estimated ROI (incremental or total project)** • **Note: ROI may not fully capture safety concerns**
DATA SECURITY	• How does risk profile change with a distributed model? • How comfortable are we with that potential risk profile? • Do we have adequate security resources to transition? • What other resources are needed?	• **Not a constant or clear relationship** • **Work closely with cyber security team to figure out impact**

CONCLUSION

AI/ML in edge computing

Shifting toward a more distributed computing infrastructure does not reduce the need for communication in multiple directions across the network altogether.

AS DATA IS PROCESSED, THE ANALYTICS PERFORMED ARE INCREASINGLY DONE USING MACHINE LEARNING-BASED TECHNIQUES AND ALGORITHMS.

As more data is generated and more subsequent decisions and results produced, these algorithms should continue to “learn” and generate better results over time, in theory at least. This continuous refinement to the algorithm requires sharing sets of data and results from multiple edge sources. Ideally, the edge sources would send back the key decisions and results to the central repository. The algorithm would be updated periodically, and an update could be pushed to the edge devices with the latest algorithm for immediate use in data processing and decision making. To summarize,

THERE IS A RISK OF ISOLATION AT THE EDGE SOURCES THAT WOULD BE AN IMPEDIMENT TO LONG TERM OPTIMAL DECISION MAKING.

Leaders should be aware of this potential roadblock and address it, if necessary.