COMPUTING PARADIGMS FOR RESEARCH:

CLOUD VS. EDGE

LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY LUT

School of Energy Systems

Degree Programme in Electrical Engineering (Industrial IoT) Master's Thesis

2022

Juha-Jaakko Heiskari

Examiners: Associate Professor, D.Sc. (Tech) Pedro Nardelli

Junior Researcher, doctoral candidate, M.Sc. (Tech) Mehar Ullah

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT Energiajärjestelmät

Sähkötekniikan Industrial IoT-koulutusohjelma

Juha-Jaakko Heiskari

COMPUTING PARADIGMS FOR RESEARCH: CLOUD VS. EDGE

Diplomityö

2022

154 sivua, 33 kuvaa, 6 taulukkoa ja 1 liite

Tarkastajat: Apulaisprofessori, TkT. Pedro Nardelli
Nuorempi tutkija, tohtorikoulutettava, DI. Mehar Ullah

Avainsanat: Microsoft Azure, Cloud computing, Edge computing, Internet of Things

Pilvipalvelut ovat muuttaneet tapaa, miten elämme, työskentelemme, ja pystymme työskentelemään jopa silloin kun jokin radikaali tapahtuma kuten Covid-19 pandemia pysäytti normaalin elämän.

Samanaikaisesti jatkuvasti kasvava Internet of Things (IoT) sekä Industrial Internet of Things (IIoT) laitteiden määrä on tuonut esiin erilaisen paradigman. Sen sijaan että tietoa koottaisiin keskitettyyn palveluun, laitteet reunalla generoivat massiivisen määrän tietoa, joka täytyy prosessoida reunalla tai hyvin lähellä sitä.

Tämän diplomityön tavoitteena on antaa lukijalle ymmärrettävä kuva Microsoft Azure -ratkaisuista, sekä siitä miten se omalta osaltaan vastaa pilvi vastaan reuna -haasteeseen. Työssä perehdytään pilven ja reunan arkkitehtuuriin, tekniikkaan ja tietoturvaan, sekä selvitetään kyselytutkimuksella, miten kohdeyrityksessä tulevaisuuden tilaan suhtaudutaan.

Työn tuloksissa selvisi, että IoT tulee kasvamaan tulevaisuudessa voimakkaasti, minkä lisäksi että reuna- ja pilvipalvelut täydentävät toisiaan, mikä johtaa uuteen tilanteeseen, hajautettuun pilveen.

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

School of Energy Systems

Degree programme in Electrical Engineering (Industrial IoT)

Juha-Jaakko Heiskari

COMPUTING PARADIGMS FOR RESEARCH: CLOUD VS. EDGE

Master’s Thesis

2022

154 pages, 33 figures, 6 tables and 1 appendix

Examiners: Associate Professor, D.Sc. (Tech) Pedro Nardelli

Junior Researcher, doctoral candidate, M.Sc. (Tech) Mehar Ullah

Keywords: Microsoft Azure, Cloud computing, Edge computing, Internet of Things

Cloud computing has changed how we live, work, and manage to keep working, even when cataclysmic events such as the Covid-19 pandemic seized everyday life.

Simultaneously, the rising amount of Internet of Things (IoT) and Industrial Internet of Things (IIoT) devices have brought in a different paradigms.

Instead of focusing on centralized computing, the distributed devices at the edge generate massive amounts of data that must be handled at or near the edge itself.

The aim of this thesis is to give the reader a comprehensive view of how Microsoft Azure solutions comply with the field of constantly developing cloud and edge paradigm.

The architecture, technology, and security aspects of both cloud and edge will be unfolded, and the current situation of platforms at the case company surveyed.

The results of this thesis show that the IoT solutions are growing exponentially, and both edge and cloud computing platforms are supplement to each other, leading to a new way to handle the paradigm, a distributed cloud.

ACKNOWLEDGEMENTS

This socially distanced thesis was accomplished for VTT Technical Research Centre of Finland Ltd. (VTT) during the Covid-19 pandemic labeled years 2021-2022.

I want to express my deepest gratitude to this master's thesis supervisors, Junior Researcher Mehar Ullah and Associate Professor Pedro Nardelli for patiently helping me out remotely throughout the research.

All the researchers, scientists, and coworkers at VTT who helped me out during my endeavor deserve my most profound gratitude. You really are doing things like the company motto "*Beyond the obvious*" states. Keep up the brilliant work!

This whole thesis would have been impossible to finish without the help of my supervisor Juha Lähde at VTT. Thank you, Juha, for endlessly supporting me on my work/school balance during this period when everything how we used to do business together changed in a blink of an eye.

I want to give special thanks to Joosua Santasalo, Microsoft Azure MVP, for his constructive criticism on the thesis.

To my family - I want to express my gratitude for having such a kind, empathetic, and maybe a bit eccentric family that I can always rely on.

Finally, hugest thanks go to my dear wife Minna for ultimate patience during my studies. Without you and your support, this thesis would never have seen daylight. Thank you for constantly reminding me there is life out there.

Juha-Jaakko Heiskari

Vantaa, April 2022

LIST OF SYMBOLS AND ABBREVIATIONS

5G	Fifth Generation mobile telecommunication system
AAA	Authentication, authorization, and accounting
AAD	Azure Active Directory
AI	Artificial Intelligence
AKS	Azure Kubernetes Service
AMQP	Advanced Message Queuing Protocol
API	Application Programming Interface
AR	Augmented Reality
ARM	Azure Resource Manager
ASA	Azure Stream Analysis
AWS	Amazon Web Services
Azure IRA	Azure IoT Reference Architecture
BI	Business Intelligence
C2D	Cloud-to-Device
CaaS	Container as a Service
CaC	Configuration as Code
CAPEX	Capital Expenditure
CC	Cloud Computing
CD	Continuos delivery
CDN	Content Delivery Network
CI	Continuous integration
CNCF	Cloud Native Computing Foundation

CoAP	Constrained Application Protocol
CPS	Cyber-Physical System
CPU	Central Processing Unit
CSP	Cloud Service Provider
CV	Connected Vehicle
D2C	Device-to-Cloud
DDoS	Distributed Denial of Service
DL	Deep Learning
DLT	Distributed Ledger Technology
DMZ	Demilitarized Zone
DNS	Domain Name Service
DPS	Device Provisioning Service
EC	Edge Computing
EDR	Endpoint Detection and Response
ETSI	European Telecommunications Standard Institute
EULA	End-user License Agreement
FaaS	Function as a Service
FC	Fog Computing
FPGA	Field-programmable gate array
FR	Face Recognition
GCP	Google Cloud Platform
GDPR	General Data Protection Regulation
HCI	Hyper-converged Infrastructure
HSM	Hardware Security Module

HTTP	Hypertext Transfer Protocol
IaaS	Infrastructure as a Service
ICS	Industrial Control System
IIC	Industrial Internet Consortium
IIoT	Industrial Internet Of Things
IIRA	Industrial Internet Reference Architecture
IoT	Internet Of Things
IP	Internet Protocol
IT	Information Technology
JEA	Just-enough-access
JIT	Just-in-time
JSON	JavaScript Object Notation
M2M	Machine-to-Machine
MCAS	Microsoft Cloud App Security
MCC	Mobile Cloud Computing
MCU	Microcontroller Unit
MDC	Micro Data Center
MEC	Mobile Edge Computing
ML	Machine Learning
MQTT	Message Queuing Telemetry Transport
NaaS	Network as a Service
NB-IoT	Narrow Band IoT
NFV	Network Function Virtualization
NIST	National Institute of Standards and Technology

OCR	Optical Character Recognition
OPC	Open Platform Communications
OPEX	Operational Expenditure
OS	Operating System
OT	Operational Technology
OTA	Over-the-Air
PaaS	Platform as a Service
PEC	Pervasive Edge Computing
PKI	Public Key Infrastructure
POC	Proof of Concept
POP	Point-of-Presence
POSIX	The Portable Operating System Interface
QoS	Quality of Service
RBAC	Role-Based Access Control
REST	Representational State Transfer
RTO	Research and technology organization
RTOS	Real-time Operating System
SaaS	Software as a Service
SASE	Secure Access Service Edge
SASL	Simple Authentication and Security Layer
SDK	Software Development Kit
SDLC	Software Development Life Cycle
SDN	Software Defined Networking
SDP	Software Defined Perimeter

SIEM	Security Information and Event Management
SLA	Service Level Agreement
SOA	Service Oriented Architecture
SOAR	Security Orchestration and Automated Response
SOC	Security Operations Center
SOM	System on Module
SSL	Secure Sockets Layer
STaaS	Storage as a Service
TCO	Total Cost of Ownership
TCP	Transmission Control Protocol
TLS	Transport Layer Security
TSI	Time Series Insights
VM	Virtual Machine
VNet	Virtual Network
VR	Virtual Reality
VTT	VTT Technical Research Center of Finland Ltd.
WSAN	Wireless Sensor and Actuator Network
XDR	Extended Detection and Response
ZTNA	Zero Trust Network Access

TABLE OF CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

Cloud computing (CC) is a well-known phenomenon in the modern-day. Before the cloud computing era, for example, when building an application, businesses needed a number of things to make it a success: Own infrastructure, safe space to house the hardware, IT employees to build, maintain, and monitor everything. All these problems are solved with cloud computing. Together with application development, cloud computing platforms have made it possible for individuals and businesses to store, process, and access massive amounts of data.

Some popular examples of everyday cloud solutions which people use are email applications, collaboration software, and messaging apps like WhatsApp. During the Covid-19 pandemic, more and more businesses and individuals rely on cloud solutions on a daily basis. Cloud enables businesses to easily develop, deploy and manage their applications for millions of users around the world.

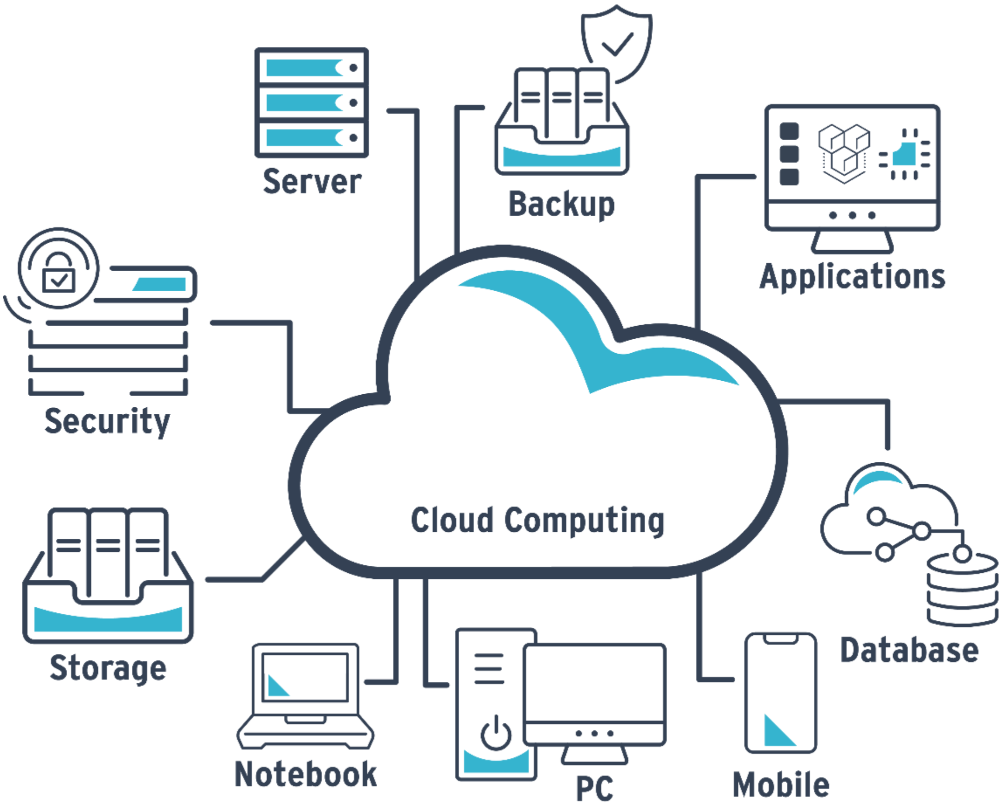


Figure 1: Use cases for cloud computing

According to Gartner, global end-user expenditure on public cloud services is estimated to hit $480 billion in 2022. The popularity of hybrid, multi-cloud, and edge systems is growing, opening the way for new distributed cloud models. Additionally, new wireless communications breakthroughs such as 5G version 16 and 17 will accelerate cloud adoption to a new level of acceptance. Gartner forecasts that public cloud investment would surpass 45% of total company IT spending by 2026, up from less than 17% now. (Gartner, Inc., 2021c).

Edge computing (EC) is a new form of computation that is performed closer to end-users rather than in data centers. Good examples are autonomous vehicles, smart thermostats, and traffic lights. In the past years, the rapidly rising amount of Internet of Things (IoT) and Industrial Internet of Things (IIoT) devices, sensors, and the massive amount of data they generate on the edge have brought new opportunities to innovate. As the demand for fast communication increases to meet this need, 5G and artificial intelligence (AI) technologies are being introduced. These new capabilities have brought along new opportunities such as autonomous decision-making in self-driving vehicles and fleet maintenance with predictive maintenance. A tiered architecture that draws from edge technologies should be embraced to meet these needs. This will bring core centralized architectures' capacity, latency, and privacy drawbacks to the fore.

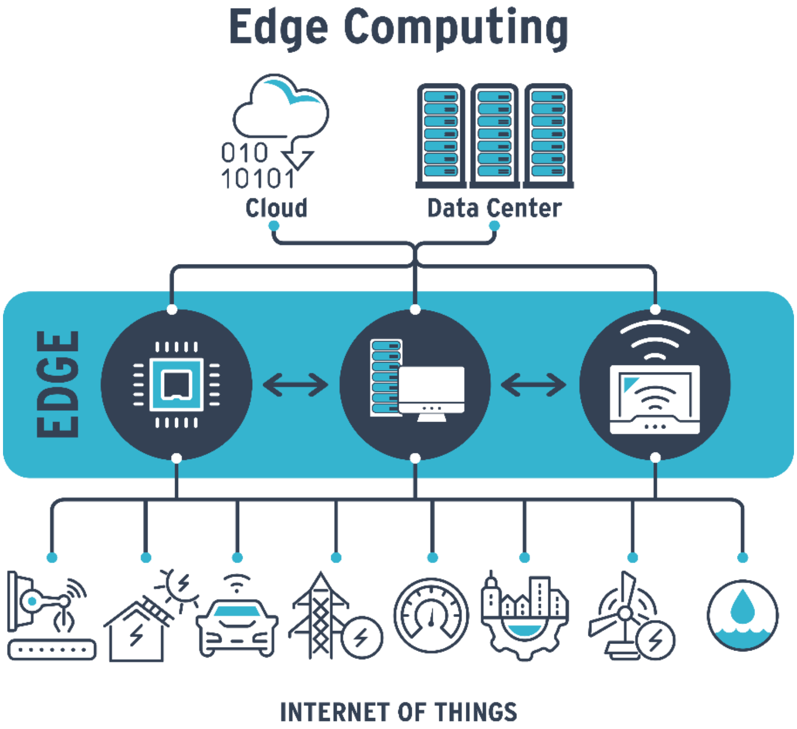


Figure 2: Use cases for edge computing (Adapted from IEEE, 2021 © IEEE).

The Internet of Things is associated with edge computing and may be seen as a reversal of the cloud computing paradigm, in which all more advanced operations take place in data centers. Instead of relying only on local assets to gather and transfer data to the cloud, decisions are made nearer where data is born, on the intelligent edge. The growth of IoT solutions is accelerating day by day, and one possible outcome of this phenomenon is a future where every device is connected.

To fully get the best out of both technologies, an entire paradigm shift is required, and to understand the paradigm shift, it is essential to know what cloud and edge computing can offer and how they differ from each other.

1.1 Background

The case study company VTT is fully Finnish state-owned and one of Europe's leading research and technology organizations (RTO) with a total revenue of 244 M€ in 2020. Through science and technology, VTT helps its clients develop new businesses and identify solutions to global concerns. VTT is an active collaborator with universities, research institutes, companies, and many European Union research programs. (VTT Technical Research Center of Finland Ltd., 2021).

Prior to closing down the thesis subject, the author had discussions with VTT's CIO, Mr. Petri Kujala, about the thesis goals, both organizational and academic.

The concept of servitization rose up several times during the discussions with Mr. Kujala. According to Kowalkowski et al. (2017), the definition of servitization is the "*transformation of a company's resource base and organizational capabilities and structures from product-centric to service-centric. It involves a redefinition of the firm's mission and routines and shared norms and values.*" Despite several instances of effective servitization in a variety of industries, literature is scarce on successful servitization in RTOs. To summarize the primary goal, the better servitization of the computing platforms is a key concept of VTT's IT roadmap to support business one step ahead.

The need for better servitization leads to the main questions for VTT researchers on the case study part of the thesis.

- Which Cloud/Edge platform(s) create the most significant impact and benefit the research the most?
- What is the role of IT tools and support in enhancing the research process in RTO?
- What IT-oriented aspects could be optimized for better performance from research and RTO's perspective?

To cover the scientific contribution of this thesis, the following research questions are surveyed.

- What does the future of cloud and edge computing look like?
- What are the primary concerns about cloud and edge computing platforms in terms of security?

Thus, basic comprehension of both paradigms is needed to understand the possible confrontation of cloud and edge platforms. The thesis uncovers these questions progressively.

1.2 Thesis goals, delimitations, and disclaimers

Because the study is theoretical in nature, descriptive research was conducted, with a literature review serving as a primary source of data. Secondary data sources include technical documentation, articles, white papers, and websites.

The work's objective is to provide the reader with a core theoretical understanding of cloud and edge paradigms, followed by a comprehensive examination of the Microsoft Azure cloud platform and how it addresses the cloud vs. edge dilemma. Additionally, a survey was conducted to determine the current state of platforms at the case study company VTT, providing data for future decision-making.

This thesis concentrates on the Microsoft Azure platform since including all the other significant players and their offerings into the equation would make this thesis bloated. Microsoft Azure is constantly developing, and the number of features is astounding. There is no point in listing every single feature of it; this thesis will not cover many exciting things to ensure the scope would be tolerable.

This thesis does not cover multiple interesting and crucial topics, including the impact of data protection legislation (e.g., GDPR, California Consumer Privacy Act) on how research can gather data, especially from data-security and data-management points of view.

This thesis is an independent publication and is neither affiliated with, nor authorized, sponsored, or approved by, Microsoft Corporation. All product names, logos, brands, trademarks, and registered trademarks mentioned in the thesis are the property of their respective owners.

1.3 Structure of the thesis

Chapter 2.1 introduces the fundamental concepts of cloud computing, from its characteristics to deployment and service models to examples of cloud computing's downsides.

Chapter 2.2 begins by defining the IoT and demonstrating its implementation. Significant technologies, building blocks, key factors, and reference architectures are all discussed.

Chapter 2.3 begins with an overview of edge computing and why it is required. Following that, a list of edge computing enabling technologies and building blocks will be provided. The thesis addresses some of the several outstanding concerns surrounding edge computing.

Chapter 2.4 summarizes the preceding chapters, starting with the relevance of cloud and edge computing from the EU's perspective and ending with the distributed cloud as a potential solution.

Chapter 2.5 is an in-depth case study of Microsoft Azure, packed with the bits and pieces that contribute to each of the thesis's paradigms.

Chapter 2.6 provides an analysis of the current cloud and edge platform threat environment. The chapter ends with demonstrating the Zero Trust concept, which enables a more secure operating environment in today's complex world.

Chapter 2.7 contains a survey-based assessment of platforms for the case company VTT.

Chapter 3 discusses the cloud vs. edge concepts and the survey findings.

Chapter 4 summarizes and finishes the thesis.

2 THESIS

This chapter provides the reader with the theoretical background of cloud, IoT, and edge computing platforms and the contrasts between them. Furthermore, this thesis digs deeper into the Microsoft Azure platform and current security threats. At the end of this chapter, the evaluation of platforms in the case company will be described.

2.1 Cloud computing platforms landscape

As data has become more essential for increasingly more tasks, humans have needed ever better ways to store and access it since the advent of information technology. Valuable information is now primarily stored on computer hard drives and central servers. These can rapidly and effortlessly store and process massive quantities of data.

Nevertheless, both hard drives and servers have their limits, necessitating businesses and industries to find a storage solution that can store and process increasing amounts of data. Companies relied on hardware and software to take their businesses online before cloud computing platforms. To reach their intended audience, businesses had to purchase these components.

Furthermore, businesses needed people who were skilled in managing hardware and software, as well as monitoring the entire infrastructure. Despite its utility, this approach had its share of difficulties, some of which were significant. There were high setup costs, complicated components, and limited storage space.

Cloud Computing was created to enable new capabilities and technologies that are accessible and affordable to everyone. And therefore make it possible to free up resources in businesses, allow people to be more efficient in their work, and empower organizations to reinvent their business models and procedures. (Bansal, 2020; Ikink, 2021).

Cloud computing, as defined by the National Institute of Standards and Technology (NIST), is: "*a model for enabling on-demand network access to a shared pool of configurable* *computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.*" (Mell and Grance 2011).

Organizations and entities that provide cloud services to consumers are called cloud service providers (CSPs). They provide consumers the facilities for running workloads and abilities to manage them, for example, with configuration portals. Today, four of the largest and most popular cloud service providers are Amazon Web Services (AWS), Microsoft Azure, Alibaba, and Google Cloud Platform (GCP). (Gartner, Inc., 2021e).

The following tables describe cloud computing essential characteristics, deployment and service models. (Adapted from Liu et al., 2011; Mell and Grance 2011; Lea, 2018; Calles, 2020; Giannoutakis et al., 2020; Intel Corporation., 2021).

Cloud Computing essential characteristics

On-demand self-service	Without dealing with the cloud provider, consumers may supply computing resources on-demand.
Broad network access	Resources are available via the network and are accessed by established standards, allowing for their use by a diverse range of devices and systems.
Resource pooling	By using a multi-tenant model, the service provider may pool its computing resources and make them available to several customers at once, all while responding dynamically to changes in customer demand.
Rapid elasticity	In certain circumstances, capabilities may be supplied and withdrawn automatically, allowing for quick expansion and contraction in response to demand. For clients, provisioning capabilities seem boundless and may be used anytime and in any way they want.
Measured service	Cloud technologies' automatic resource management and optimization are achieved by implementing a metering capability at the appropriate abstraction level for the type of service. Monitoring and reporting resource utilization gives better transparency for service providers and customers.
Cost model	Instead of the traditional model where the consumer has to buy the appropriate hardware and focus on capital expenditure (CAPEX), cloud

	computing services operate on a pay-as-you-go approach that focuses on operational expenditure (OPEX).
Reliability	Cloud service providers handle underlying service-related technicalities with signed Service Level Agreements (SLAs).
Security	Cloud service providers provide highly secure storage for their customers where data is kept secure and private so it cannot be tampered with. They also offer advanced security features, which are more reliable and practical than traditional in-house systems not designed for modern environments.

Table 1: Cloud computing essential characteristics (Adapted from Liu et al., 2011; Mell and Grance 2011; Lea, 2018; Calles, 2020; Giannoutakis et al., 2020).

Cloud Deployment Models

In the cloud environment, three forms of cloud topologies are commonly used: private, public, and hybrid cloud. Additionally to the primary three, community and multi-cloud are introduced. All cloud frameworks should be capable of dynamic scaling, development, and deployment, as well as the appearance of locality independent of proximity or model.

Private clouds can mean components that are handled on-premise. Modern corporate systems often use a hybrid architecture to safeguard mission-critical applications and data on-premise while using the public cloud for connectivity, simplicity of deployment, and quick development.

Private cloud	Infrastructure is provided only for one business or organization in a private cloud. There is no concept of resource sharing or pooling outside of the owner's own infrastructure. Within the premises, sharing and pooling are everyday activities. A private cloud exists for several reasons, such as security and assurance, to guarantee that customer-managed systems are the only ones that have access to customer-managed data. However, cloud-like qualities such as virtualization and load balancing must be present for it to be deemed a cloud. A private cloud may be on-premises, or it can be dedicated hardware supplied only for their use by a third party. It can also be located off-site.

Public cloud	In comparison to a private cloud, a public cloud is a total opposite. The infrastructure is made available on a demand basis to a diverse range of clients and applications. Infrastructure is a pool of resources that everyone may use at any moment as part of their service-level agreements. The benefit here is that the sheer scale of cloud data centers enables extraordinary scalability for a large number of consumers, who are only limited by their desire to purchase a certain amount of the service. The cloud infrastructure is built for public use. It may be a corporation, an academic institution, a government agency, or a mix of these organizations that own, manage, and operate it. It resides on the service provider's facilities.
Hybrid cloud	Two or more public, private, or community clouds are combined in the hybrid architecture concept. These combinations may involve the concurrent usage of numerous public clouds or a mix of public, community, and private cloud infrastructure. If sensitive data needs specific management, companies choose a hybrid solution, which leverages the cloud's reach and scalability via the frontend interface. Another scenario is to have a public cloud agreement in place to cover for instances when the corporation's private cloud infrastructure is unavailable. The public cloud will act as a load balancer in this case until the flow of data and usage returns to the private cloud's limits. Cloud bursting is a concept that refers to the practice of using clouds as a source of contingent resource provisioning. The hybrid cloud configurations are linked through standardized or proprietary technology that enables the mobility of apps and data.
Community cloud	The infrastructure of the community cloud is supplied only for the use of a select range of clients from comparable businesses. One or more community organizations, a third party, or a mix of the two may own, manage, and operate the infrastructure. It might be on- or off-site.
Multi-cloud	A multi-cloud architecture makes use of resources from different providers, which enables it to host a massive number of applications. A federated cloud is also a kind of multi-cloud environment. A multi-cloud

	strategy requires a comprehensive design overview in order to accommodate all necessary cloud providers.

Table 2: Cloud computing deployment models (Adapted from Liu et al., 2011; Mell and Grance 2011; Lea, 2018; Calles, 2020; Giannoutakis et al., 2020).

Cloud Service Models

Each service model of cloud computing is designed to address a distinct set of business needs.

NaaS (Network as a Service)	Software-Defined Networking (SDN) and Software-Defined Perimeters (SDP) are examples of NaaS services. These solutions are cloud-managed and structured techniques for overlay network and corporate security provisioning. Instead of developing a worldwide infrastructure and capital to support a company's communications, a cloud technique may be employed to construct a virtual network. This enables the network to scale resources up or down in response to demand appropriately, and new network capabilities may be swiftly acquired and implemented.
STaaS (Storage as a Service)	Cloud storage that a client rents from a service provider is termed StaaS, or Storage as a Service. Corporations, businesses of all sizes, and individuals may all utilize the cloud to store data and perform data backup and recovery. The primary advantage of STaaS is that it enables the expense and work associated with managing data storage infrastructure to be outsourced to a third-party service provider. This significantly simplifies scaling storage capacities without requiring the acquisition of additional hardware or incurring setup expenditures. Additionally, clients can react more quickly to shifting market circumstances. With a few clicks, a client may rent terabytes or more of storage without requiring to set up new storage gear on the client's end.
IaaS	IaaS was the first proposal of a cloud service. The supplier delivers scalable hardware services in the cloud and provides a specific set of

(Infrastructure as a Service)	software frameworks for consumers to install and use any software, including operating systems and apps. To do this, the consumer must provide more lifting on their end.
PaaS (Platform as a Service)	PaaS is an acronym for a platform as a service, which refers to the cloud's underlying hardware and lower-layer software services. In this case, the end-user utilizes the data center hardware, operating system, middleware, and numerous frameworks provided by the provider to host their own application or service. Middleware may include database systems. The advantage of a PaaS deployment over an IaaS deployment is that the client benefits from the scalability and low operating costs associated with cloud infrastructure while also using established middleware and operating systems from the provider. Because the bulk of the componentry, operating system, and middleware are assured to be available, and provided the client's overall application adheres to the vendor's framework and infrastructure requirements, the customer benefits from a speedier time to market.
SaaS (Software as a Service)	The foundation of cloud computing is SaaS. Typically, a provider offers applications or services to end-users through clients such as mobile devices, thin clients, or frameworks on other cloud platforms. On the user side, the SaaS layer is almost entirely contained inside their client. This software abstraction has facilitated the industry's rapid expansion in the cloud. SaaS solutions include Google Apps, Salesforce, and Microsoft Office 365. In a word, SaaS enables consumers to access apps hosted on a provider's cloud infrastructure and accessible through a client interface.
CaaS (Container as a Service)	Software container creation and orchestration tools such as Docker and Kubernetes are provided to customers as a cloud computing service called Container as a Service (CaaS). It allows customers to assemble all the software components required for an app into a single container without requiring the client having to create the infrastructure.

FaaS (Function as a Service)	This type of service offering allows a customer to execute separate software functions and link them together to construct an application. FaaS (also called serverless computing) enables a customer to lease the computer time necessary to complete activities without maintaining any software or hardware. FaaS offers the same benefits as PaaS and CaaS, but without the requirement for customers to set up platforms and containers. FaaS was first presented as a public cloud service, but as the product progressed, FaaS is currently capable of supporting all cloud deployment methods.

Table 3: Cloud computing service models (Adapted from Liu et al., 2011; Mell and Grance 2011; Lea, 2018; Calles, 2020; Giannoutakis et al., 2020; Intel Corporation., 2021).

Cloud Computing's Drawbacks

Vendor Lock-In

When comparing the benefits and drawbacks of cloud computing, one of the most significant problems is vendor lock-in. Many companies claim that using the cloud and integrating their business needs with the cloud service provider is an easy process; however, leaving them to another vendor can be quite challenging. This observation applies to both apps and platforms. The transition could be risky and may be impractical because of support difficulties. (Alalawi and Al-Omary, 2020).

Cost is always a factor in system design, and the cloud may provide both cost savings and resource ease. Customers must adequately grasp cloud pricing systems, billing cycles, service level agreements (SLAs), and cloud capabilities prior to migrating their applications to the cloud. Otherwise, the choice may not result in cost savings, and the tenant may get stuck in a costly SLA with a vendor. (Dhirani et al., 2017).

Different cloud suppliers offer a variety of pricing schemes that are established on a coarse- or fine-grained structure. For example, the types of instances, such as general-purpose or storage-optimized, and their availability, are determined by the tenant's decision to reserve or rent them. Each instance has unique storage, processing, memory, and availability characteristics, and these characteristics will determine which instance is the best match for the customer's needs. (Dhirani et al., 2017).

Adopting a multi-cloud strategy and designing flexibility with future portability in mind, assisted with a thorough understanding of cloud pricing mechanisms are some of the well-known ways to avoid dependence on one vendor.

Downtime

According to Alalawi and Al-Omary (2020), when customers experience downtime, which could lead to data loss and lost revenue, they can become highly dissatisfied. There is no way to avoid downtime because all cloud computing systems are internet-based. If users are in a location without internet access, they will be unable to cloud-based data, software, or apps. Because outages can occur regardless of whether users expect them or not, no organization is immune to downtime. For instance, on March 3, 2020, Microsoft's Azure cloud was down for six hours at the US East data center, limiting the availability of Azure cloud services for many North American customers according to Tsidulko (2020). It became one of the most talked-about cloud outage topics of 2020. Some techniques for minimizing downtime include implementing a disaster recovery strategy with the shortest recovery time possible and designing a multi-region deployment.

Limitations of control and adaptability

Businesses have minimal control over their data since cloud service providers own and administer the whole cloud infrastructure. Because end-user license agreements (EULAs) and management policies vary across service companies, they must adhere to the provider's specifications. The usual situation for this configuration is that it allows the client minimal control and limits their access to cloud-based apps, tools, and data. Due to their absence, the client may be unable to access critical administrative services. In the worst-case situation, they limit customers' ability to use their implementations. To maintain flexibility and control, best practices include understanding the customer's and cloud provider's responsibilities in the contract, SLA, and support level. (Alalawi and Al-Omary, 2020).

Security

Security in cloud services is not always an easy topic to understand. When it comes to storing enterprise data, cloud computing offers some pros and cons. Cloud-based data storage makes it susceptible to cyber-attacks. By 2025, however, a Gartner research predicts that consumer mistakes would account for 99 percent of data breaches and cyberattacks. (Gartner, Inc., 2019). Security issues will be covered in further detail in Chapter 2.6.

In summary, cloud computing allows users to do computer operations using physically distant resources. Cloud computing's almost infinite capacity for storage, memory, and processing units allows it to augment the computational capabilities of IoT devices and components. However, this fusion of cloud and edge poses considerable challenges, including increasing complexity, as detailed in the following chapters.

2.2 Defining IoT

The IoT is a vital first step in establishing a worldwide infrastructure that connects machines and people. It offers significant social and economic potential across sectors such as agriculture, mining, health care, manufacturing, and transportation. However, it also creates substantial security, safety, and technological infrastructure delivery and management challenges.Consumer services and enterprises dominated the Internet revolution's first two decades, but they were human-centric. Businesses have been profoundly impacted by business-to-business models and the cloud, wiping out whole sectors that failed to adapt quickly enough to the revolution's fast speed. (Serpanos and Wolf, 2018).

Pervasive information systems, sensor networks, and embedded computers are all examples of IoT technologies with historical antecedents. The major portion of IoT devices are linked in order to form purpose-built solutions. (Serpanos and Wolf, 2018).

Sensors are hardware components that detect and quantify a physical attribute. The word "things" refers to all of these sophisticated sensors and gadgets. Not only can these Things detect and measure physical properties, but they also record and store an infinite quantity of data on a network. Due to evolving technology, these Things can analyze this data utilizing embedded intelligence. (Bansal, 2020).

IoT has exploded in popularity over the previous decade. However, it represents various things to different individuals. According to Whitmore et al. (2015, as referenced in Lynn et al., 2020) and Serpanos and Wolf (2018), there has been no consensus on the word's single, universal meaning so far.

Whitmore et al. (2015, as cited in Lynn et al., 2020) define the IoT as "*a paradigm where everyday objects can be equipped with identifying, sensing, networking and processing capabilities that will allow them to communicate with one another and with other devices and services over the Internet to achieve some objective*."

According to Shin (2014, as referenced in Lynn et al., 2020), loT is founded on a number of essential principles and enabling technologies. Examples are the identification of objects (things), information sensing, data sharing technologies, and network integration technologies. The IoT was not included in the legacy computer and telecommunications infrastructure architecture.

New computing paradigms are necessary because of the sheer number and variety of devices, and also the extraordinary volume, diversity, and speed of data generated by them. According to Lynn et al. (2020), "*depending on the use case and service level requirements, IoT devices may require processing and storage locally, in the cloud, or somewhere in between.*"

According to Gartner, "*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*." Decentralizing computing power or shifting it closer to the data generation source — in other words, seeking edge computing — might be helpful. This progress is fueled by the fast deployment of IoT projects for various government, corporate, and consumer use cases, such as smart cities. Gartner estimates that "*around ten percent of enterprise-generated data is currently produced and handled outside of a centralized data center or cloud.*" Gartner forecasts that this percentage will reach 75% by 2025. (Gartner, Inc., 2018).

Examples of IoT Systems

IoT systems are beneficial in a broad variety of domains. Sensors, for example, may be used to monitor industrial processes and the condition of the equipment. A growing number of electric motors have sensors that alert the user about approaching motor failure. Sensors in smart buildings determine the number and position of individuals as well as the building's status. Other examples include monitoring pedestrian and vehicular traffic with smart city sensors. Autonomous cars are outfitted with a network of sensors that continuously monitor the vehicle and its surroundings. Medical systems enable the connection of a diverse array of patient monitoring sensors. (Serpanos and Wolf, 2018; Mijuskovic et al., 2021).

IoT Implementation status

According to a recent survey conducted by Hypothesis Group and Microsoft (2021), 90% of respondents, who were IoT decision-makers from different businesses worldwide, have adopted IoT, which is still being utilized for a variety of objectives that result in enhanced efficiency and productivity. The survey underlines that "*those who use IoT for cloud security, supply chain management, and sustainability more strongly believe IoT is critical for their success.*"

The majority of firms responded are embracing AI, edge computing, and digital twin initiatives. About 80 percent of them strive to incorporate the technology into their IoT solution. Regardless of widespread use, many initiatives remain in the trial or Proof of Concept (PoC) phase mainly to "*a lack of infrastructure and the complexity of scaling and managing systems*," the survey reveals. In the future, enterprises will need professional support since the degree of implementation differs by the industry for "*certain technologies. E.g., Smart Places is advanced in AI, Energy is advanced in Edge Computing, and Manufacturing is advanced in Digital Twins,*" according to the survey by Hypothesis Group and Microsoft (2021).

Definitions of key technologies in IoT

The following table describes the definitions of key technologies in IoT. (Adapted from Mell and Grance, 2011; Wang and Leblanc, 2016; Iorga et al., 2018; Yousefpour et al., 2019).

Cloud computing	A concept for providing a shared pool of customizable resources and services that may be swiftly supplied and released with little administrative effort or participation from service providers.
Dew computing	Dew computing is a cloud computing concept for on-premises computer software-hardware organizations. The on-premises computer serves many purposes in collaboration with cloud service providers, independently of cloud services.
Edge computing	Edge computing refers to the network layer that surrounds end devices and their users, enabling sensors, meters, and other network-connected devices to do local computations.
Fog computing	Multilayered fog computing architecture enables ubiquitous access to shared, scalable computing resources. It enables latency-aware distributed applications and services. It is made up of physical or virtual fog nodes that are located between intelligent end devices and centralized cloud services.
Mist computing	Mist computing refers to distributed computing at the edge, namely the IoT devices themselves. It was conceived with the goal of enabling future self-aware and autonomous systems.

Table 4: Definitions of key technologies in IoT (Adapted from Mell and Grance, 2011; Wang and Leblanc, 2016; Iorga et al., 2018; Yousefpour et al., 2019).

IoT Building Blocks

Understanding IoT's building blocks is crucial to comprehending its functioning and relevance; these parts operate in tandem to provide its functionality. According to a study by Ullah et al. (2020), IoT comprises six essential components that interact together to enable its functionality:

The identification block is used to identify network devices uniquely. The device's name and address in the communication network are known as the object ID and the object address, which are used to identify devices. IPv6 and IPv4 are the two primary protocols for addressing IoT things.

Sensors acquire data about things and surroundings inside the communication network and transfer it to the target database or cloud. Cloud-based analytics are used to examine the acquired data. Actuators, which are physical devices such as switches that operate in the opposite direction of sensors, are also employed in IoT systems. This category is called **the sensing block.**

Numerous disparate items communicate with one another and with the platform in order to share data and services. In **the communication block**, according to the study, "*IoT communication protocols such as Message Queue Telemetry Transport (MQTT) and Constrained Application Protocol (CoAP) are used to connect different items to the IoT and transmit data from those connected objects to the management system*." BLE, Wi-Fi, NFC, and ZigBee are examples of technologies used to connect devices to the Internet.

The computation block is composed of both hardware and software components. Numerous hardware platforms, including Arduino and Raspberry Pi, have been utilized to operate IoT apps. Similarly, IoT features are implemented using a number of software platforms. The operating system serves as the device's primary software platform and is active during the activation time. According to the study, the cloud platform is also a computing component of the IoT; it allows the transmission of data from small objects to the cloud, real-time processing of big data, and assisting the end-user in extracting information from it.

The services block: IoT services assist developers of IoT applications by offering a point to begin for development. When developers are aware of existing services, they are more concentrated on developing the IoT app than constructing the infrastructure and services that will support it. Four distinct types of IoT service categories exist:

Identity-related services are categorized as either active or passive. **Active identity-related services** broadcast data and are powered by a continuous power supply or battery. Active identity-related services have the capability to transfer or transmit data to another device. **Passive identity-related services** are not powered externally and must be transmitted through external devices or mechanisms. Passive identity-related services are limited to reading data from devices. According to the study, the procedures of acquiring data from sensors, processing it, and providing it to an IoT app for processing are referred to as **information aggregation services. Collaborative aware services** make judgments and respond appropriately based on the data offered by information aggregation services. **Ubiquitous services** make collaboratively aware services available to everyone, everywhere, and anytime.

The semantic block may be called the IoT's brain because it is central to the process of knowledge extraction, which involves locating and utilizing resources, modeling information, assessing data, and selecting and providing the suitable service. (Ullah et al., 2020).

Key Factors of an IoT Platform

According to the study by Ullah et al. (2020), IoT system's central component is an IoT platform. The market for IoT platforms is flooded with hundreds of vendors, making it challenging to locate and pick a trustworthy and scalable IoT platform. Businesses may be able to identify and pick the optimal IoT platform for their operations if they take certain essential variables into account prior to making a platform selection choice. Platform needs are context-dependent, and platforms do not have to include all of the criteria of the study mentioned below, although they may.

Stability: The market is flooded with hundreds of platforms, each with its own set of issues. Some systems may be incapable of providing services to clients. As a consequence, a platform with a high probability of market survival should be picked.

Scalability and flexibility: In the beginning, a firm may be tiny, and its business sector is limited. However, as the business expands, so does the business sector. Scalability to business needs is critical to ensuring that the platform can serve the company throughout its evolution. Due to the quick evolution of current technology and market needs, the platform's technology should be adaptable.

The pricing model and business case: Certain vendors give an introductory discount during the first few months but subsequently significantly raise the price. Additionally, some vendors provide a great discount to entice users, but the contract contains just a limited number of services, and adding extra ones is highly expensive. As a consequence, it is essential to choose a platform that provides all the features a company needs at a price range that matches its budget.

Security is critical in the IoT since all platforms must be of high quality. The study state that security may include addressing issues such as "*device-to-cloud network security, data encryption, application authentication, secure session initiation, app authentication, cloud security, and device security*."

Time-to-market: Platform providers' support and time-to-market throughout the product lifecycle, from idea through sales, should be considered. Additionally, certain vendors offer customized startup packages to new clients, which assist in several stages of product development.

Data analytics and visualization tools: Prior to deciding on an IoT platform, it is critical to ascertain the required data analysis and visualization demands. The most capable platform for data collecting, processing, and visualization capabilities should be identified.

Data ownership is a tricky topic with the Internet of Things. Different jurisdictions have varying legal systems and interpretations of the law. For instance, the European Union has different data ownership standards and rules than the United States. As a result, it is vital for the supplier of an IoT platform to comprehend data rights and data protection's geographical scope.

Ownership of cloud infrastructure is expensive, and some minor IoT platform vendors focus just on the software layer. Certain providers' platforms have been certified by one or more of the largest public cloud providers; for that reason, the majority of their solutions are delivered through a single leading platform. It is necessary to evaluate the larger business cloud's interoperability with the IoT platform provider.

The extent of legacy architecture: Because IoT devices may communicate with a range of different infrastructure platforms, it is frequently unclear how an existing IoT system is connected. As a result, organizations and enterprises should evaluate how the interaction between new and older technology work when implementing an IoT platform.

Protocols such as Hypertext Transfer Protocol (HTTP), Advanced Message Queuing Protocol (AMQP), and MQTT are required for IoT systems to function. MQTT is very lightweight and has far fewer overheads due to its binary nature. As a consequence of technical breakthroughs, new devices are being introduced to the market. The IoT platform chosen should be capable of supporting new protocols and should be easily upgradeable.

System performance: When an event happens in an IoT platform, a rule-based trigger may be executed automatically. Due to the fact that IoT solutions offer such rule-based triggers, the average time required to assess and manage each event rises as the number of connected devices grows. Before choosing an IoT platform, it is necessary to analyze the provider's efforts to guarantee that the platform operates appropriately.

Interoperability: The Internet of Things platform is a middleware solution. Numerous programs will use the obtained data, which may or may not be accessible through the platform. As a consequence, the IoT platform of choice should be open-source compliant. The company's efficiency will grow as a result of interoperability.

Redundancy and disaster recovery: IT infrastructure difficulties may strike at any moment, and in the case of IT infrastructure failures, IoT platform providers should have a dedicated backup infrastructure in place. The timeline for data backups and the availability of failover clusters on the IoT platform should be considered.

Attractive interface: To guarantee that customers can easily use the IoT platform's features, its interface should be clear, appealing, and convenient.

Application environment: Prior to deciding an IoT platform, three features of the application environment should be considered: the apps that come pre-installed, the application development environment's characteristics, and the standard interfaces. Certain IoT platforms may be integrated with on-premises IT systems. A hybrid cloud may be advantageous in certain instances since business-sensitive or mission-critical functions may be conducted locally. Simultaneously, the platform may handle public and less sensitive processes.

Platform migration: As a business expands, the IoT platform may find itself unable to handle all of its needs. As a consequence, the need for a more established IoT platform supplier may arise. Consequently, organizations should verify that the IoT platform provider they choose has well-documented interfaces, schema, and Application Programming Interface (API) in order to facilitate potential migration to other IoT platforms.

Prior experience: Before making a decision, a business needs to ascertain if the IoT platform vendor has worked on comparable projects to the business application. A lengthy history of working in the same area is a positive indicator.

Bandwidth: Low latency and high bandwidth networking are required for the IoT platform to provide optimal information and communication between processing components. As a result, it is critical to verify that a prospective platform vendor has sufficiently broad network lines and expansion capabilities.

Edge intelligence and control: Distributed, offline, and edge intelligence are the future of IoT systems. When devices can make judgments based on local data rather than waiting on cloud decisions, they become more powerful. As a consequence, the IoT platform should be developed in such a way that it can use edge intelligence and support new topologies. (Ullah et al., 2020).

Industrial IoT

IoT's application to the industrial sector is called the Industrial Internet of Things (IIoT). IIoT is defined by Boyes et al. (2018, as cited in Lynn et al., 2020) as: "*A system comprising networked smart objects, cyber-physical assets, associated generic information technologies and optional cloud or edge computing platforms, which enable real-time, intelligent, and autonomous access, collection, analysis, communications, and exchange of process, product and/or service information, within the industrial environment, to optimize overall production value.*"

In IIoT, numerous (smart) industrial machines, actuators, and sensors link with one another to create a network of intelligent IoT-enabled devices. Industrial applications may benefit from IIoT solutions, according to Narayanan et al. (2020), by "*improved connectivity, efficiency, profits, scalability, and data speeds for industrial applications*," thus improving safety, predictive maintenance, and operational efficiencies. However, the IIoT's maximum potential can be realized only via the implementation of a flexible communication system that can suit a variety of needs. (Narayanan et al., 2020).

IoT Reference Architectures

Reference architectures enable IoT software developers to comprehend, compare, and assess multiple IoT systems reliably. Numerous reference architectures have been suggested to standardize the principles and implementation of IoT systems across multiple domains, with two of them being briefly introduced in this study.

Utilizing the **Industrial Internet Reference Architecture** (IIRA), businesses may gather and analyze data from physical items or other massive data streams using sensors, software, machine-to-machine (M2M) learning, and other technologies. (Lynn et al., 2020).

The IIRA is a standards-based open architecture for IIoT systems. The IIRA optimizes its value by being adaptable to a wide range of industries in order to facilitate interoperability, map relevant technologies, and guide technology and standard development. To achieve wide industry application, the architectural description and representation are generic in nature and at a high degree of abstraction. The IIRA abstracts and distills common traits, features, and patterns from Industrial Internet Consortium (IIC) and other use cases. The IIRA will be enhanced and changed in response to suggestions gained from its usage in IIC testbeds and real-world deployment of IIoT systems. By identifying technological gaps based on architectural needs, the IIRA design aspires to transcend today's technologies. This will fuel the efforts of the industrial internet community to build new technologies. (Industrial Internet Consortium, 2019).

The **Azure IoT Reference Architecture** (Azure IRA) connects sensors to intelligent cloud services through the Microsoft Azure platform. The Azure IRA’s fundamental objective is to act on business insights obtained by IoT application data collection. (Microsoft, 2021b).

The reference study presented by Lynn et al. (2020) recommends an IoT architecture, describing fundamental principles and ideas, IoT subsystem specifics, and considerations for solution design. Azure IRA was designed with versatility in mind. As a consequence, it generates IoT solutions that are based on microservices architecture and are cloud-native by nature. It is proposed that scaling, upgrading individual IoT subsystems, and allowing for technology selection freedom per IoT subsystem are all advantages since deployable services are self-contained.

Figure 3 illustrates the Azure IRA that is suggested for integrating hybrid cloud and edge solutions. According to the study, the user interface, stream processing, and IoT devices are all represented by the color green. Green also denotes an IoT gateway. There must be communication between the IoT device and the cloud gateway that manages the device. The stream processor consumes, stores, and integrates data into the business process. The Azure IRA proposes a particular technology for each subsystem that leverages Azure services. Subsystems such as user administration, machine learning, data transformation, and IoT edge devices are available as options. On-site aggregation, transformation, and data processing are possible with edge devices, whereas cloud-based data transformation may change and convert telemetry data. The IoT system can operate appropriately by learning from previous data using the machine learning subsystem. Lastly, the user management subsystem equips users with the necessary tools for device administration. (Lynn et al., 2020).

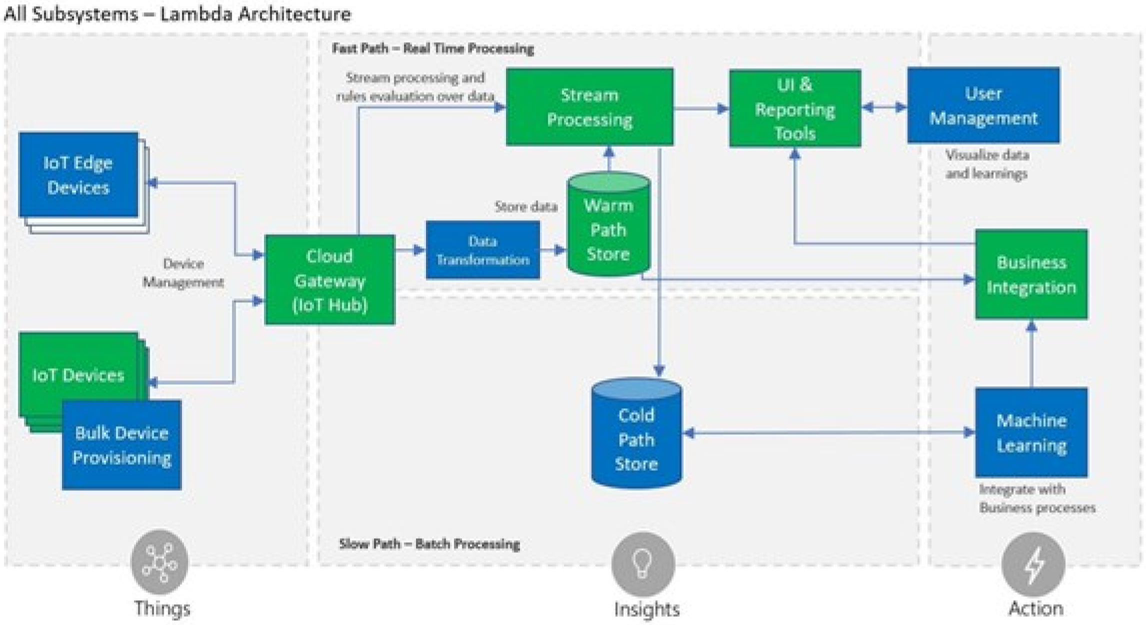


Figure 3: Azure IoT Reference Architecture (Used with permission, Microsoft, 2021b).

2.3 Edge computing platforms landscape

Edge computing definitions

Edge computing definitions include a wide range of statements. For example, Gartner defines edge computing as "*a part of a distributed computing topology in which information processing is located close to the edge, where things and people produce or consume that information.*" (Gartner, Inc., 2021a).

Another description by Red Hat's senior principal marketing manager, Rosa Guntrip, is "*Edge computing refers to the concept of bringing computing services closer to service consumers or data sources. Fueled by emerging use cases like IoT, augmented reality (AR) / virtual reality (VR), robotics, machine learning, and telecommunications network functions requiring service provisioning closer to users, edge computing helps solve the critical challenges of bandwidth, latency, resiliency, and data sovereignty. It complements the hybrid computing model where centralized computing can be used for compute-intensive workloads while edge computing helps address the requirements of workloads that require processing in near real-time.*" (Overby, 2021).

The COVID-19 outbreak has triggered a rethinking of cloud tactics. According to Gartner, "*collaboration, mobility, and virtual desktops are rapidly shifting to the cloud to enable a remote and secure workforce.*" Simultaneously, Gartner reports that the number of IoT devices doubles every five years, posing serious security risks. Cloud computing use is increasing as business applications move to the public cloud and enterprises adopt cloud-native architectures. The usage of edge computing is increasing as hyper-scale cloud providers develop strategies for deploying cloud capabilities closer to the edge. Edge computing addresses a growing need for low latency, more data processing at the edge, and network resilience in the case of a network failure. (Gartner, Inc., 2021b).

Edge computing may help real-time systems such as actuator and sensor networks, transportation, and industrial automation by attempting to reduce latency, offer location awareness, and enhance the Quality of Service (QoS). The extensive geographical distribution of edge computing is one of its distinctive characteristics. At the edge, services are supplied through devices like set-top boxes or access points. Edge computing aims to bring computational resources and services from the cloud closer to the user. (Ahmed et al., 2017).

According to Ikink (2021), edge computing may address latency, bandwidth, and security concerns, as the Internet of Things, Container-as-a-Service, and AI are among the fastest expanding cloud services. Edge computing enables enterprises with privacy and compliance issues to achieve a greater degree of information security.

Edge data centers, cloudlets, and fog computing have been created in response to the fact that cloud computing is not necessarily efficient at processing data produced at the network's edge. (Cao, Zhang, and Shi, 2018).

Due to the fact that the data created by IoT devices is kept and processed inside edge network nodes, an edge computing platform may help solve privacy issues. Additionally, offloading computation to resources located near to consumers and the data centers close by them might assist in minimizing energy usage at end devices. (Premsankar, Di Francesco, and Taleb, 2018).

According to IDG Communications, Inc. (2021), at its most fundamental level, edge computing enables processing and data storage to be moved closer to the users rather than depending on a centralized site thousands of kilometers away.

Any networking or computing resource that exists between cloud data centers and data sources is referred to as "Edge". For instance, a smartphone, a smart home gateway, a Cloudlet or a Micro Data Center (MDC) serve as an edge between the cloud and the destination. Figure 4 shows the two-way compute streams in Edge computing paradigm, where objects serve as both data consumers and producers. Not only may entities at the edge request services and data from the cloud, but they can also execute computations from the cloud. The edge may offload computation, data storage, caching, and processing from the cloud to the user, as well as distribute cloud-based request and delivery services to the user. The edge itself must be well-designed in order to effectively satisfy requirements for services such as dependability, security, and privacy protection. Edge computing will have as significant an impact on our society as Cloud computing. (Cao, Zhang, and Shi, 2018).

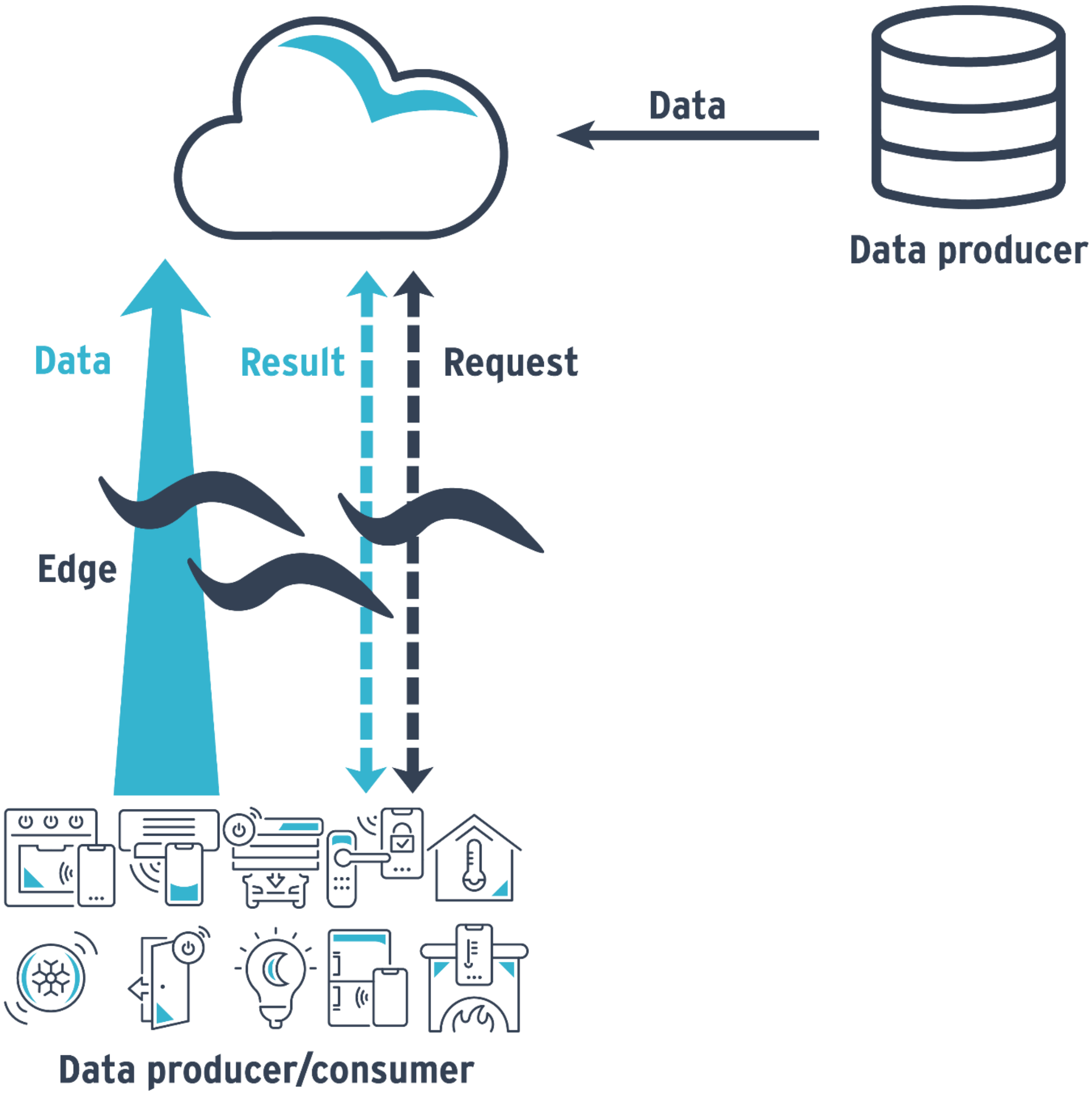


Figure 4: Edge computing paradigm (Adapted, with permission, from Cao, Zhang, and Shi, 2018 © Springer).

Edge computing business model

According to Cao, Zhang, and Shi (2018), cloud computing's economic model is pretty straightforward. Users directly buy services from service providers depending on their needs. Cloud services are often delivered through dynamic, virtualized, and easily expandable Internet resources. These services might include information technology infrastructure, software, and other Internet-related services or resources. Edge computing applies to various disciplines, including information technology, communication technology, and a variety of industrial linkages. Edge computing's economic model should be driven not just by customer demand for services but also by data.

The enterprise's stakeholders decide the business model for edge computing. One of the open issues is how to combine the present Cloud computing business model with a multilateral Edge computing business model. (Cao, Zhang, and Shi, 2018).

Why is edge computing necessary?

Cloud Services Push

Placing all computing jobs in the cloud is an efficient method of data processing because cloud computing power significantly exceeds the capabilities of edge devices. As the volume of data created at the edge increases, the speed at which data is sent becomes a gridlock for the Cloud computing concept. If all information had to be sent to the cloud for processing, response times would be unacceptably lengthy. By supporting several edge nodes in a single network segment, current network capacity and dependability would be overstressed. Data must be handled at the edge in this situation to provide quicker response times, more efficient processing, and less network load. (Cao, Zhang, and Shi, 2018).

Internet of Things Pull

Over time, all sorts of electrical equipment linked to the Internet will all be included in the IoT, serving as both data providers and consumers. It is reasonable to predict that the number of items at the edge of the network will exceed billions in a few years. They will create a massive volume of raw data, rendering conventional Cloud computing incapable of handling it all. As a result, the vast majority of IoT data will never be sent to the cloud. Rather than that, it will be consumed at the perimeters of the network. (Cao, Zhang, and Shi, 2018).

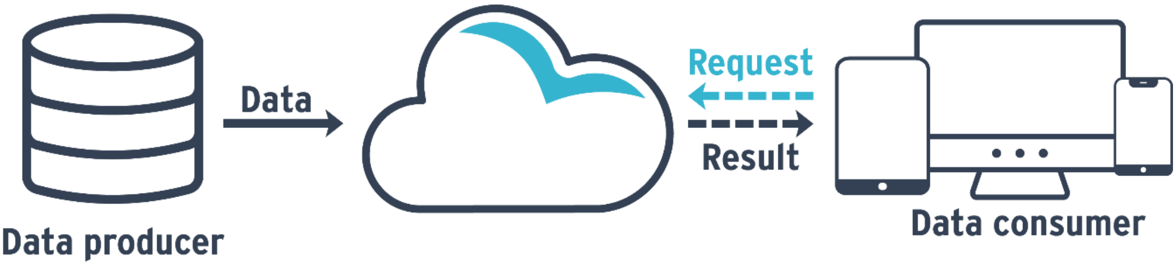


Figure 5: Cloud computing paradigm (Adapted, with permission, from Cao, Zhang, and Shi, 2018 © Springer).

The traditional cloud computing design is seen in Figure 5. As demonstrated by the solid black line, data producers create raw data and upload it to the cloud, while data consumers submit data consumption requests to the cloud. The blue dotted line represents the data consumption request submitted to the cloud by data consumers, while the black dotted line depicts the outcome. However, this structure is inadequate for IoT. To begin, the data volume at the edge is enormous, necessitating an inordinate amount of bandwidth and computational resources. Second, the demand for privacy protection will impede Cloud computing in IoT. Lastly, the majority of IoT end nodes are energy-constrained. Due to the high energy consumption of the wireless connection module, moving some processing operations to the edge may be more energy efficient. (Cao, Zhang, and Shi, 2018).

The transition from Data Consumer to Data Producer

The Cloud computing paradigm's edge devices are often data consumers, such as a smartphone playing a YouTube video. On the other hand, individuals are increasingly creating data through their mobile devices. The journey from being a data consumer to being a data creator and consumer needs a greater emphasis on edge function placement. For instance, it is relatively common for individuals to record videos or shoot photographs and then share them via a cloud-based platform such as TikTok or Instagram. However, the picture or video clip may be huge, necessitating a significant amount of bandwidth for uploading. Before uploading the video clip to the cloud, it should be converted to a suitable resolution at the edge. Second example is wearable health gadgets. Because the physical data acquired by devices at the network's edge is often private, processing it locally may be more efficient than sending raw data to the cloud. (Cao, Zhang, and Shi, 2018).

Third example is the smart grid technology idea which has arisen in the energy industry as a way to boost the efficiency and flexibility of the conventional grid. Essentially, a smart grid is an electrical network comprised of infrastructure, software, and hardware. It enables two-way communication between the generating and consumption sides via the use of energy-efficient equipment such as power electronics converters, energy storage systems, smart appliances, smart meters, as well as other energy-efficient resources. Edge computing offers substantial advantages to organizations that are enabled by the IoT. Cloud computing, on the other hand, is still crucial since numerous advantages accrue from centralizing data storage and analysis. (Ullah et al., 2021).

Key technologies that enable edge computing

By definition, **virtualization** is "*a process that abstracts operating systems, computer resources, storage devices, and network resources*," according to Cao, Zhang, and Shi (2018). More precisely, computation virtualization refers to the process of building a virtual machine (VM) that mimics the behavior of a physical computer equipped with an operating system (OS). (Cao, Zhang, and Shi, 2018).

VMs have been extensively implemented in data centers and even on standalone servers. Virtualization enables the simultaneous operation of several VMs on a single physical server's CPU. Virtualization enables programs to be segregated across VMs and adds a layer of protection by preventing the data of one application from being easily accessible to another. Virtualization optimizes resource use on a physical server, increases scalability due to the ease of adding or upgrading applications, and decreases hardware costs, among other benefits. Each VM is a complete computer that runs all components, including its own OS, on top of virtualized hardware. (Kubernetes, 2021b).

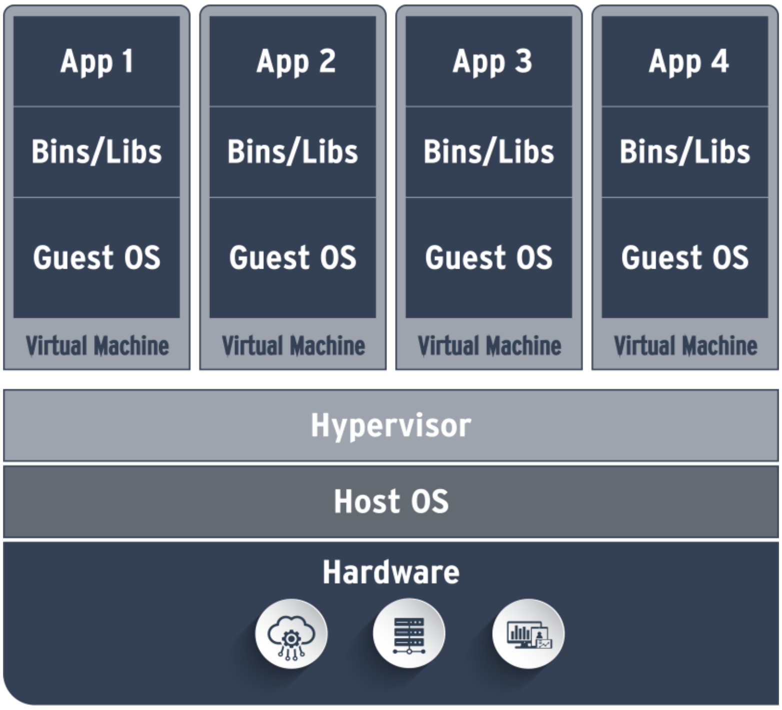


Figure 6: Hypervisor-based virtualization stack (Adapted from Kubernetes, 2021b).

As seen in Figure 6, four VMs run on the same hardware device utilizing an OS-level emulation, each of which contains a guest OS, runtime support, and the program. This results in a substantial waste of resources. However, an OS emulation capability enables the hypervisor to run many operating systems on the same bare metal.

When shipping containers were developed in the 1950s, they changed the global economy. **Containerization** is making a comeback, but this time on the cloud. Containerization is the process of enclosing a program and all of its dependencies in a standardized set of lightweight libraries and APIs. It is a standardized method for storing and shipping all components, guaranteeing the seamless operation of an application. Gartner estimates that 70% of worldwide enterprises will have more than two containerized apps in production by 2023, as cited by Ikink (2021).

Containers are comparable to VMs, but they have less strict isolation features, enabling programs to share the OS. This technique enables the creation of numerous segregated user-space instances. A program running on a standard operating system has access to all of the computer's resources (attached devices, network shares, and CPU power). However, apps operating within a container have access to just the container's contents and devices. The size of the containers may be limited to a few megabytes, and startup time may be as little as a few seconds. The portability of containers is ideal for Edge computing applications, where resource constraints, such as storage space and reaction time, are often constrained. (Cao, Zhang, and Shi, 2018).

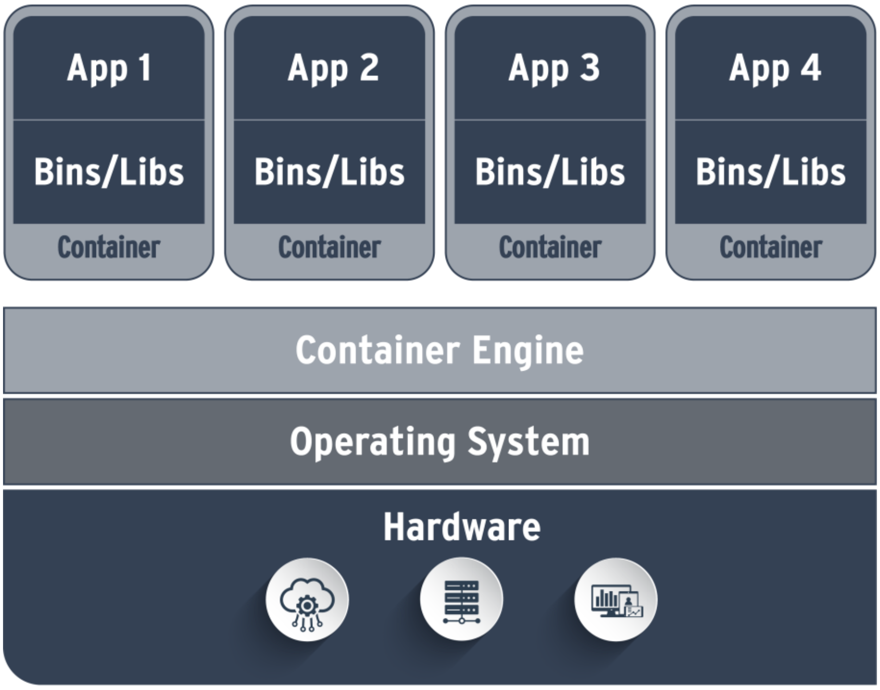


Figure 7: Container-based virtualization stack (Adapted from Kubernetes, 2021b).

As seen in Figure 7, container-based virtualization does not include a hypervisor or a guest OS. No guest OS is required since container instances encapsulate the application and its dependencies at the process level. A container image contains a base image, which is a customized version of the operating system, programs, and dependencies like libraries and binaries, and, most often, a series of instructions for configuring and starting the application. On the host OS, the container operates as an isolated user-space process, whereas all containers share the OS kernel. (Cao, Zhang, and Shi, 2018).

According to Cao, Zhang, and Shi (2018), "*the reason why virtualization is vital in edge computing is the data abstraction.*" It allows data to be processed at or near its origin rather than being transported to a centralized location such as a data center. Data abstraction is not a priority in cloud computing since all data is sent to centralized data centers where a distinct data abstraction occurs. However, with edge computing, the volume of data is too vast to be sent to the cloud, therefore data must be wisely aggregated and routed to different applications, where compute and networking virtualization can support such data abstraction at the edge devices and servers. (Cao, Zhang, and Shi, 2018).

Serverless computing architectural solutions enable start-ups and small enterprises to scale rapidly without investing capital. Between 2020 and 2025, this demand is predicted to boost the serverless sector by 25%. Event-driven architectures operate in response to particular events, requiring little to no human intervention in the event of an emergency. (Ikink, 2021).

Network virtualization is a revolutionary method that simplifies network administration by allowing for the programmatic setup of networks and the abstraction of core network operations. Software-defined networking (SDN) and network function virtualization (NFV) are the two main concepts. SDN aims to decouple network control and forwarding operations, whereas NFV aims to abstract forwarding and other networking tasks from the underlying hardware. (Cao, Zhang, and Shi, 2018).

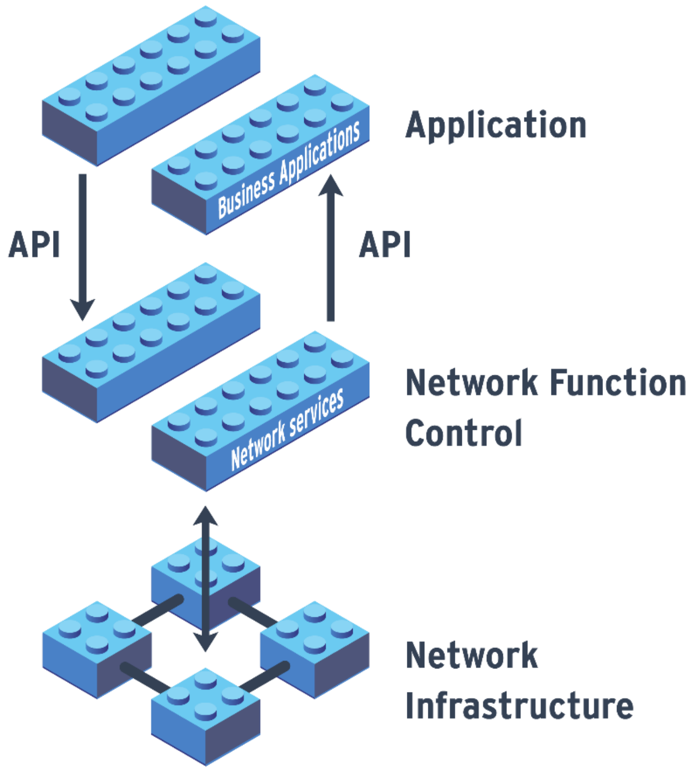


Figure 8: A high-level architecture overview of SDN (Adapted, with permission, from Cao, Zhang, and Shi, 2018 © Springer).

SDN is a dynamic, manageable, and cost-effective architecture. It enables the abstraction of the underlying infrastructure from applications and network services and the configuration and deployment of edge devices in a plug-and-play fashion. Additionally, SDN is a viable approach for securing Edge systems like IoT, smart cities and homes. The three levels of SDN are seen in Figure 8. The control plane is comprised of one or more controllers that act as the SDN network's brain. The controller layer maintains a global view of the network, which appears as a single logical switch to applications and policy engines. The centralization of intelligence has limitations in terms of security, scalability, and flexibility, which is the main problem with SDN. (Cao, Zhang, and Shi, 2018).

NFV is the process of implementing network functions into software modules that can operate on general-purpose hardware. NFV provides the autonomous deployment of virtual resources to handle an unexpected spike in traffic caused by an IoT application at a given location. In conjunction with NFV, SDN allows the flexible and programmable deployment of software-based modules, simplifying network design and maintenance. (Premsankar, Di Francesco and Taleb, 2018).

According to Pan and McElhannon (2018), the key enabling technologies for the future edge cloud, such as NFV and SDN, are still in their infancy. For edge new application delivery, there are no established standards yet.

Initially, for Edge computing, no consideration was made for **Content Delivery Network**s (CDNs). However, caching content on Edge servers located near data users fits the Edge computing philosophy rather nicely. As the server that streams the material becomes the web's gridlock as a result of increased web traffic, CDNs may enable scalable data caching at the network's edge, significantly reducing bandwidth costs and page load times. (Cao, Zhang, and Shi, 2018).

Micro Data Centers (MDC) and Cloudlets are a new type of highly mobile small-scale cloud data centers. They may be used to connect Edge/mobile devices to the cloud. Due to their geographical location, the Edge devices could access the computational capacity on the MDCs or Cloudlets with decreased latency. To decrease resource costs, important computing functions like "*speech recognition, language processing, machine learning, image processing, and augmented reality (AR) may be hosted on Cloudlets or MDC,*" according to Cao, Zhang, and Shi (2018).

Edge computing building blocks

At first, the physical architecture of edge computing might seem perplexing. However, it consists mainly of client devices communicating with a nearby edge module in order to perform responsive and fluid functions. These modules are considered critical components of the infrastructure of edge computing. Numerous systems are built to use edge computing hardware and services as a source of processing and storage on the local level. To decrease bandwidth needs, edge gateways process data and transmit just what is required. If a real-time application is necessary, it may communicate back with the edge device. IoT sensors, laptop computers, cellphones, security cameras, and even internet-connected domestic appliances are all examples of edge devices. (IDG Communications, Inc., 2021).

Cloud computing allows individuals to create their own code and then distribute it to the cloud. The cloud provider is responsible for determining the computing location in the cloud. Users have no or just a rudimentary understanding of how the program works. This is one of the advantages of cloud computing: the user is unaware of the infrastructure. Typically, since the application operates only in the cloud, it is often built to run on a specific platform and developed in a single programming language. (Cao, Zhang, and Shi, 2018).

On the other hand, edge computing offloads computation from the cloud, and the edge nodes are almost always heterogeneous platforms. These nodes' runtime varies, and the programmer has significant challenges while developing an application for deployment in the Edge computing paradigm. (Cao, Zhang, and Shi, 2018).

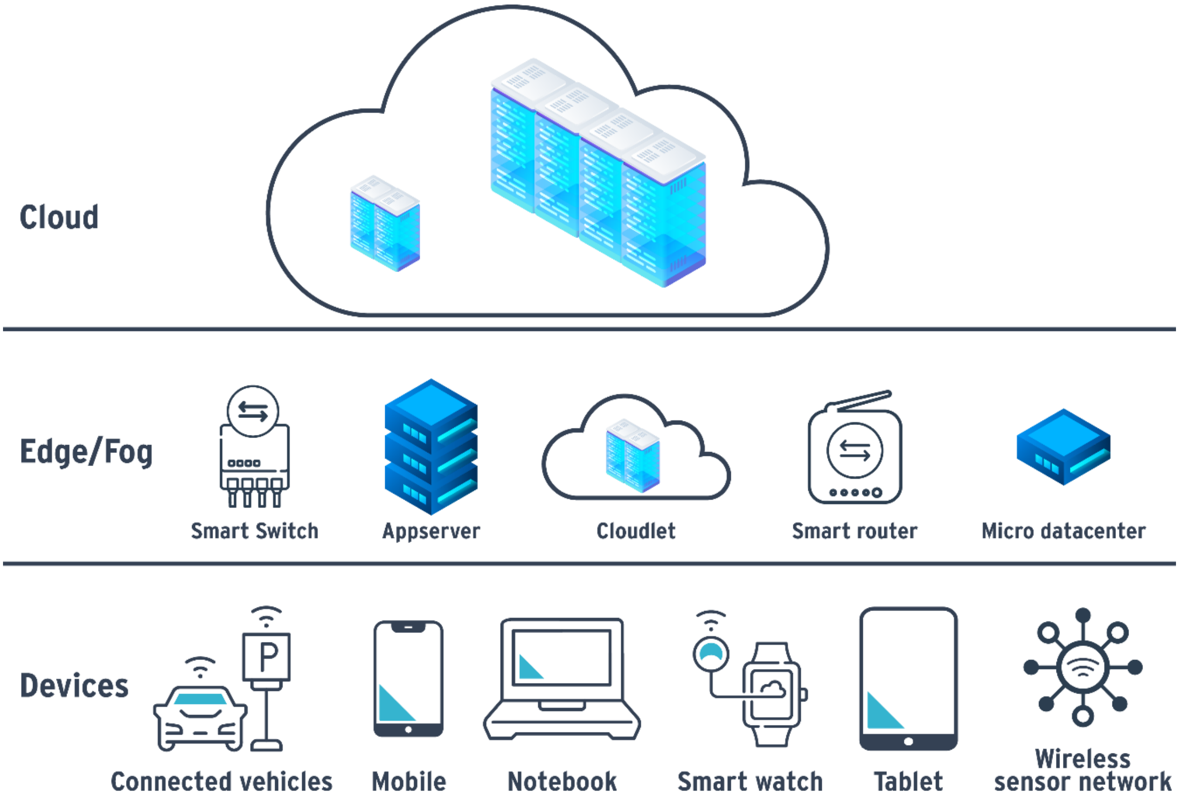


Figure 9: A high-level hardware hierarchy of the edge computing paradigm (Adapted, with permission, from Cao, Zhang, and Shi, 2018 © Springer).

Figure 9 depicts a high-level overview of the edge computing paradigm's hardware hierarchy. The cloud continues to be critical at the top of this architecture for centralized data storage, processing, and analysis. Edge computing, in the center, serves as a regional hub for geographically distributed applications. A substantial share of the work required to process data is performed by a variety of different kinds of hardware, such as a micro data center, a router, or a base station router. Edge devices, at the base level, serve as bridges connecting edge services and end-users. Any device with sensing capabilities that collects data from users and the surrounding environment, such as a mobile phone or a laptop, qualifies as an edge device. It is worth mentioning that there is no apparent distinction between edge devices and edge servers, with some overlap. (Cao, Zhang, and Shi, 2018).

Given the hierarchy, a provider's job as a supplier of infrastructure and capabilities that enable others to harness edge computing resources will include resource virtualization, provisioning, and maintaining these resources, as well as making them transparent to users. Meanwhile, what matters most for an edge application developer is which edge devices their applications can utilize, which tools or data processing platforms are most suited for their apps, and how their apps are developed, tested, and deployed in edge computing. (Cao, Zhang, and Shi, 2018).

Edge Resource Management

Edge computing resources include not just typical computing resources such as network, storage, memory, and CPU, but also those non-traditional computing resources like energy. For instance, a smartphone's or sensor's battery life is limited, and users do not want to charge up these items constantly. Some sensors may operate for years without being recharged. Thus, while developing edge analytics apps that use data from such devices, developers must consider a variety of things, including the amount of power available to their app and the frequency with which the apps should read from or write data to the sensors. These will be a critical limitation on scheduling or task allocation optimization. In summary, according to Cao, Zhang and Shi (2018), the "*three key concerns of the resource management in edge computing are: a high degree of heterogeneity of hardware/software, dynamic availability and mobility of hardware/software, and existing tools are relatively heavy since they are usually designed for servers and clusters.*" (Cao, Zhang, and Shi, 2018).

Operating systems, virtual machine hypervisors, and specialized resource management systems are examples of conventional resource management technologies. Typically, the operating system is the most obvious requirement for the majority of edge devices, not just traditional operating systems such as Linux, but also embedded operating systems for SoCs. (Cao, Zhang, and Shi, 2018).

As previously stated in the thesis, VM hypervisors pool resources for many virtual machines/applications. In a similar fashion to VM hypervisors, two fairly recent technologies, Kubernetes and Docker, can be used to manage numerous containers on a single host or cluster. In edge computing, the majority of services and applications would be containerized, making it easy to design and deploy edge analytics across many edge devices. Moreover, containers have lower overhead than VMs due to their process-level virtualization, which makes them more ideal for edge computing, considering that the majority of devices are resource-constrained. (Cao, Zhang, and Shi, 2018).

Kubernetes and Docker

Kubernetes is a Google-sponsored open-source project that automates containerized applications' deployment, scaling, and management. The diagram in Figure 10 illustrates how physical resources and containers are managed by Kubernetes.

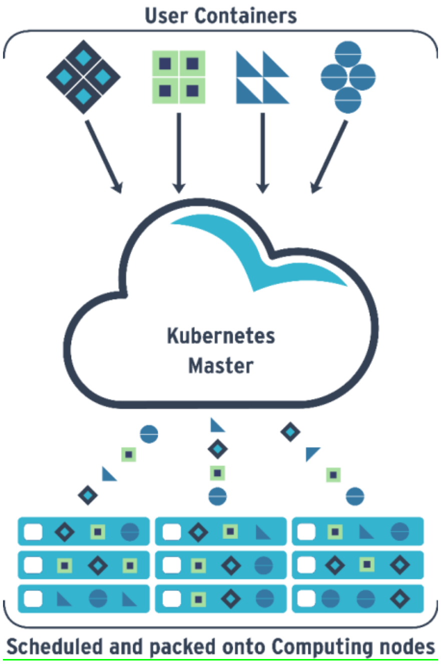


Figure 10: A high-level architecture overview of Kubernetes (Adapted, with permission, from Cao, Zhang, and Shi, 2018 © Springer).

The Kubernetes operates as a centralized scheduler that manages the placement of various containers on available resources. Another content management system that is comparable to Kubernetes in terms of functionality is Docker.

According to Kubernetes (2021b), some of Kubernetes' distinguishing characteristics are as follows:

- **Load balancing and service discovery:** Kubernetes containers may be exposed through their DNS name or their own IP address. If a container receives much traffic, Kubernetes is capable of load balancing and distributing the network traffic to ensure the deployment remains stable.
- **Storage orchestration:** Kubernetes enables to seamlessly mount any storage system available.
- **Automated rollouts and rollbacks:** Kubernetes allows to specify the intended state for the deployed containers, and it will transform the actual state to the desired state at a regulated pace. For instance, Kubernetes may be used to automate the creation of new containers for the deployment, the deletion of current containers, and the migration of their resources to the new container.
- **Automatic bin packing:** Supply Kubernetes with a cluster of nodes on which containerized jobs may be executed and specify how much CPU and memory each container requires. Kubernetes may load containers onto the nodes to optimize resource use. It is quite probable that one edge device may host numerous services concurrently in edge computing. Thus, it is critical to intelligently assign various services to available edge devices in order to obtain, for example, the greatest performance or save the most energy, depending on the use case.
- **Self-healing:** Kubernetes restarts failing containers, replaces them with new ones, destroys containers that do not react to a user-defined health check, and does not broadcast them to clients until they are ready to serve. Because many edge devices are mobile, network connectivity might be unreliable. Services operating on these devices must be restored or relocated in this situation.
- **Secret and configuration management:** Kubernetes enables the storage and management of sensitive data, including passwords, OAuth tokens, and SSH keys. Without rebuilding the container images and exposing secrets in the stack settings, secrets and application configuration may be deployed and changed. (Kubernetes, 2021b).

Purchasing edge computing service or building it in-house?

There are several ways to acquire and install an edge system. On the one hand, a firm may want to manage as much of the process as feasible internally. This would include choosing edge devices, most likely from hardware manufacturers, architecting a network that meets the use case's requirements, and purchasing management and analytic software capable of performing the required functions. That is a significant amount of effort and would need significant in-house IT experience. However, it may be an appealing alternative for a big corporation seeking a completely tailored edge deployment. (IDG Communications, Inc., 2021).

On the other hand, providers in specific sectors are increasingly selling the cutting-edge services they manage. All major cloud providers, including AWS, Azure, and GCP, provide IoT products and services. If an organization chooses this route, it may be as simple as making a contract with a vendor to install its own equipment, software, and networking infrastructure and pay a monthly charge for usage and maintenance. This has the benefit of being easily accessible and generally painless to set up, but intensively managed services may not be appropriate for many use cases. (IDG Communications, Inc., 2021).

In addition to commercial systems from cloud service providers, OpenStack, Apache Edgent, and EdgeX Foundry are just a few of the open-source edge computing solutions available. Microsoft's solution, Azure IoT Edge, is used to connect cloud-based analytics to edge devices. Three components make up the Azure IoT Edge: IoT Edge modules, IoT Edge runtime, and a cloud-based user interface. Cloud logic handles communications and activities on IoT Edge devices through the IoT Edge runtime. Numerous IoT Edge modules may be launched as Docker compatible containers on the Edge device to conduct Microsoft Azure services, third-party features, or custom code. Users may distribute and monitor workloads on IoT Edge devices using the cloud interface. (Cao, Zhang, and Shi, 2018).

This thesis will focus more on Azure IoT and Edge services in Chapter 2.5.2.

Edge computing security and privacy

Locally collected data must be managed carefully for security concerns since it is susceptible to compromise at the network's edges if not adequately protected. Remote data access has gained in popularity as more individuals work from home. When there are more remote access chances, cybercriminals have a better possibility of obtaining and abusing corporate material. Edge computing filters and processes data locally rather than delivering data to a central processing center. Businesses and their consumers will be safer if less sensitive data is sent between these devices and the cloud. (Cao, Zhang, and Shi, 2018).

Data security and user privacy are critical services that must be offered at the edge of the network. In an IoT-enabled house, a great deal of private information may be gleaned from sensed usage metrics. For instance, based on the reading of the energy or water use, one may deduce whether or not the property is unoccupied. In this scenario, the problem is to maintain service without jeopardizing privacy. Certain private information may be removed from data prior to processing, like erasing all the facial features in a video. Experts believe installing computers at the data resource's edge, i.e., in the house, might be a useful technique of protecting the privacy and data security. (Cao, Zhang, and Shi, 2018).

According to Cao, Zhang, and Shi (2018), numerous issues exist to safeguard data security and user privacy at the network's edge. Those include:

- **Lack of community understanding about privacy and security**, as seen by open WiFi networks, routers configured with default passwords, and encryption algorithms that are simple to break.
- **Ownership of the data acquired by Edge-connected devices**. As with mobile apps, the data acquired from end-users by objects will be saved and analyzed on the service provider's side. Keeping data where it is acquired, at the edge, and allowing total control to the user, on the other hand, will provide a more secure approach for privacy protection. As with health record data, end-user data acquired at the network's edge should be retained at the network's edge, with the user having discretion over how service providers use the data. Additionally, extremely confidential data may be erased during the authorization process to safeguard user privacy further.
- There is a **scarcity of practical solutions for ensuring security and data privacy at the edge of the network**. Owing to the resource constraints of some devices, the

present security protection techniques may be unable to be installed on them. Additionally, the significantly changing environment at the network's edge exposes or unprotects the network. (Cao, Zhang, and Shi, 2018).

Edge computing and Edge AI

Until recently, cognitive activities could be performed solely by massive data centers. Over time, AI found its way into software, where predictive algorithms altered the way these systems enhance organizational performance. Edge AI is a real and increasing phenomenon that powers everything from smartphones and smart speakers to vehicle sensors and security cameras. Nvidia has identified a need for more processing at the edge and is developing a variety of components that include AI capability. Nvidia's newest module, for example, is the Jetson Xavier NX, which is compact enough to be integrated into smaller devices. Due to the high computing needs of AI algorithms, the majority of these implementations make use of cloud services. This expansion of AI chipsets will result in faster reaction times for applications that need rapid computing on edge devices. (IDG Communications, Inc., 2021).

Edge Analytics

A data processing flow that uses edge computing to fully or partly execute complicated applications and services is called edge analytics. For instance, augmented and virtual reality are widely used in gaming and healthcare. Edge computing addresses the fundamental need for low response latency by processing data close to the source. Another example is health care, which includes fall detection and continuous vital sign monitoring for elderly persons. The home gateway may gather and evaluate these data and then, if required, send an alarm to a physician. (Cao, Zhang, and Shi, 2018).

Other applications like autonomous robots, field monitoring, and smart cities and homes may also benefit from edge computing. The IoT is a network of physical objects that are linked and exchange data, such as cars, household appliances, and other commodities. Although many modern technology gadgets use cloud computing, developers of IoT applications are beginning to see the advantages of doing additional computation and analytics directly on the devices. This strategy reduces latency for mission-critical apps, decreases reliance on the cloud, and improves data management amid the vast data flood created by the IoT. (Cao, Zhang, and Shi, 2018).

A portion of the data acquired by edge devices must be sent to the cloud for sophisticated, intelligent analysis. Aggregating and transferring data to the cloud or downstream apps is critical when dealing with millions of sensors or edge devices in a single application. As a result, the message hub acts as a conduit for data queuing between the edge and the cloud. Numerous queuing systems exist, including Kafka, ActiveMQ, and RabbitMQ. Cloud computing makes extensive use of these techniques. AMQP and MQTT are two critical queuing protocols that may be utilized for edge computing with or without modification. (Cao, Zhang, and Shi, 2018).

To summarize the advantages of edge computing, beginning with minimal bandwidth requirements: decisions are made at the network's very edge, requiring very little traffic, corresponding with a better end-user application experience. Because data is not sent to the cloud and back, the likelihood of intercepting data decreases, and security improves.

To highlight the risks associated with edge computing, beginning with control and contractual concerns, edge computing systems' decentralized and heterogeneous nature raises concerns about interoperability and vendor lock-in. The same qualities also highlight the difficulty of edge computing's governance and security.

2.4 Differences – and future of cloud and edge computing platforms

According to a recent Eclipse Foundation poll, businesses are growing aware of the advantages of edge computing. To stay competitive, solution vendors must have an edge computing strategy. According to the study findings, the top three operational issues for IoT and edge computing are end-to-end monitoring and administration of IoT solutions, device management, and security.

IoT technologies are gaining traction at a breakneck speed. 47% of respondents already use IoT solutions, and 39% want to do so during the next 12 to 24 months. Additionally, edge computing is gaining prominence. 54% of businesses will utilize or intend to utilize edge computing technology during the next 12 months. Additionally, 30% of businesses want to explore edge installations within the next 12 to 24 months. (Eclipse Foundation, 2021).

The importance of cloud and edge from the EU's perspective

European Commission's science and knowledge service, The Joint Research Centre, has published a paper that reviewed the EU Green Deal and its impact on IoT and Edge computing.

Several essential topics rose up, such as:

- The EU must safeguard its digital technology sovereignty.
- The pervasiveness of digital technologies necessitates a collaborative approach via international collaboration with other democratic countries.
- Protecting Science, Technology, and Innovation systems that may be disrupted in the following years may be critical to obtaining digital sovereignty.
- The EU has sovereignty over its own data.
- Cyberwarfare poses a danger to the EU's essential infrastructures.

The paper discussed the following technology developments and potential for open strategic autonomy in the future:

Increased competition in the digital technology sector creates chances for the EU to secure its digital technology sovereignty. This might be accomplished by growing up the EU's start-up ecosystem.

Due to the pervasiveness of digital technologies in our lives, a coordinated approach to their regulation is essential via international collaboration with like-minded governments. Regulatory frameworks might be created to foster innovation while adhering to EU ideals and norms aimed at enhancing the EU's leadership capability.

Protecting Science, Technology, and Innovation systems that may be disrupted in the future, developing international partnerships, and guaranteeing the competitiveness of researchers and educational institutions may be crucial for securing future digital sovereignty.

Technological sovereignty has evolved to advance the concept of European leadership and open strategic autonomy in the digital domain. Current strengths and weaknesses are inextricably linked to future possibilities and threats..

Edge computing can offer secure and distributed solutions at or near the edge of a computing network. Cyber-security concerns emerge because the number of IoT devices doubles every five years. A European Alliance on Industrial Data, Edge, and Cloud will facilitate the growth of Joint investments in cross-border cloud infrastructures and services such as European marketplaces for cloud and EU Cloud Rulebook for cloud services. Cloud computing and edge computing may help the European Green Deal achieve its environmental objectives. Along with the legislation, the EU must invest in edge computing. (Cagnin et al., 2021).

In addition to the Joint Research Center paper, a report by AIOTI, a multi-stakeholder consortium, highlights the key lessons for a more coherent and robust strategy for IoT/IIoT and edge computing at the European level.

The discovered themes and patterns have been condensed into four key proposals:

- "*Europe must build on its strengths in electronic control systems, safety-critical systems, sensing and automation, mechatronics and microelectronics/microsystems, privacy-preserving technologies, and intelligent connectivity.*
- *There is a need for a single market for IoT/IIoT edge computing founded on open standards.*

- *Europe needs a trustworthy infrastructure that builds on flexible federation and a fair business offer to manage vast amount of IoT-generated data and change how ownership and location of data are treated. The EU needs to identify the catalysts that may speed up innovation at the edge.*
- *Europe needs to capitalise on the shift of value creation to the edge. It can do this by further accelerating the technological developments of IoT and edge computing*."

(Alliance for Internet of Things Innovation, 2020).

Cloud and edge, better together, leading to a new paradigm?

According to Orrin and Chehreh (2020), edge and cloud computing are two different technologies. While traditional cloud computing is used to handle data that is not time-sensitive, edge computing is used to handle data that is. Edge computing is superior to cloud computing in places with restricted access to central processing because it avoids delay by executing computation closer to the consumers. This is an ideal match for edge computing since it allows local data processing and storage.

Edge computing advantages may also be used for highly specialized and intelligent devices. Even with a basic use case, such as the autonomous car that uses edge computing to brake before colliding with a pedestrian, constructing and distributing the edge node is only the first step. To fully realize these technologies' potential, a fundamental change is necessary. (Orrin and Chehreh, 2020).

Cloud computing is at the core of a modern operating strategy that relies on dispersed DevOps teams to create new apps. They are typically cut off from the infrastructure team responsible for cloud architecture management. Each mission's objective is no longer to install additional physical edge nodes. Rather than that, new apps will be created to match shifting demand on current nodes. (Orrin and Chehreh, 2020).

Security must be included in cloud-native systems and apps from the start. On cloud-native apps, legacy security techniques focused on perimeter protection are a formula for catastrophe. As security dangers rise, security measures should be updated. Risk, like storage, memory, computing power, and bandwidth, becomes one of the application's variables when it migrates to the edge in a cloud-native world. (Orrin and Chehreh, 2020).

The primary distinctions between IoT, edge, and cloud computing

According to Yu et al. (2018), data processing location is the distinction between what distinguishes cloud computing from edge computing. The term cloud computing refers to the central processing and storage of information in a data center, while on the other hand, edge computing refers to processing occurring at or near the source.

According to Offin (2020), emerging technologies like 5G, AR, IoT, and wearables have resulted in a flood of data created near to the user or at the network's edge. Remote working has exacerbated the problem, as an increasing number of devices seek to connect to business networks from places other than central offices. Due to the impact on network capacity, an alternative technique, such as edge computing, is necessary.

Offin's second argument is that processing data at the edge decreases cloud computational demand; organizations may employ edge computing to address this issue. However, doing so may require a total revamp of IT infrastructure. Recent advances in edge computing have rendered the classic "delete and recreate" methodology obsolete. By merging edge data centers and edge computing, enterprises can more effectively address their unique business requirements while also relieving the cloud of its own workload. (Offin, 2020).

Remote collaboration through video meetings such as Microsoft Teams during Covid-19 is an example of real-time connection, where latency concerns may impair meeting quality. Data is sent to a data center, processed, and returned to a network device. (Offin, 2020).

An example of edge computing where real-time connection concerns are more of a mission-critical issue may be a field worker using AR smart glasses to guide a physical world procedure. Latency difficulties might considerably delay their job and perhaps result in the process failing. As an edge computing solution, data processing is brought closer to the devices, lowering latency which might result in communication disruptions. (Offin, 2020).

	IoT	Edge	Cloud
Deployment	Distributed	Distributed	Centralized
Components	Physical devices	Edge nodes	Virtual resources
Computational	Limited	Limited	Unlimited
Storage	Small	Limited	Unlimited
Response Time	N/A	Fast	Slow
Big data	Source	Process	Process

Table 5: Characteristics of Iot, edge and cloud computing (Reprinted, with permission, from Yu et al., 2018 © IEEE).

IoT and edge computing are both undergoing tremendous evolution on their own. Despite their autonomy, edge computing platforms may assist IoT in resolving several crucial difficulties and enhancing performance. Thus, it has become evident in recent years that they should be incorporated. As seen in Table 5, IoT and edge computing have a lot in common. Additionally, cloud computing is used as a reference. (Yu et al., 2018).

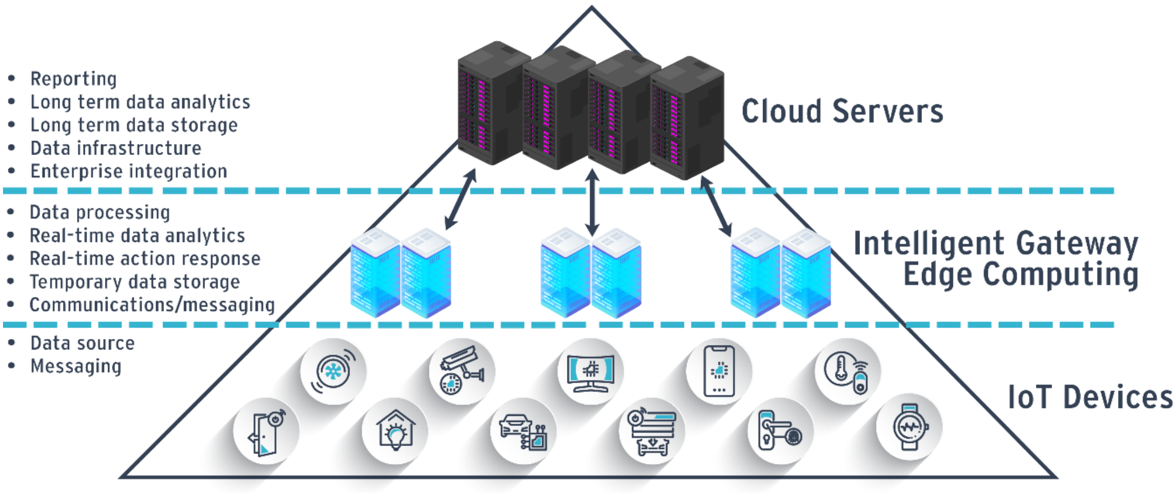


Figure 11: Three-layer architecture of edge-computing-based IoT (Adapted, with permission, from Yu et al., 2018 © IEEE).

Figure 11 shows the three-tier architecture of IoT based on edge computing. It is composed of the same layers as edge computing, all IoT devices being edge computing end users. Due to the properties of the two topologies, the IoT may generally benefit from both Cloud computing and Edge computing. Despite its limited processing and storage capabilities, edge computing offers more benefits for IoT than cloud computing. The IoT needs a rapid response rather than requiring a large amount of storage and processing capacity. Edge computing provides enough compute capability, storage, and a quick reaction time to suit the needs of IoT applications. (Yu et al., 2018).

Edge computing, on the other hand, may benefit from IoT by expanding its structure to accommodate distributed and dynamic edge computing nodes. Edge nodes may be utilized to deliver services through IoT devices or residual compute power. Numerous research projects have sought to use cloud computing to benefit the IoT, although edge computing often outperforms cloud computing in terms of competitive performance. Since the amount of IoT devices grows, edge computing and IoT are expected to become more interdependent. The majority of IoT demands may be categorized as transmission, storage, or computing. (Yu et al., 2018).

The shift to Distributed Cloud

Emerging technologies which enable distributed cloud

According to Giannoutakis et al. (2020), "*the number of connected devices is growing exponentially, with dozens of billions of things expected to come online over the next several years.*" Combining and integrating more sensors onto each device, resource management becomes more complicated. By linking these to the Internet, a massive volume of information is created in previously unheard-of numbers, diversity, and velocity, termed Big Data by El-Seoud et al. (2017, as cited in Giannoutakis et al., 2020). This information has been consolidated and is now kept on the cloud. Transferring data is very costly, much more so in vast quantities, and significantly decreases computing efficiency. The majority of cloud infrastructures extend horizontally over numerous nodes in a data center or across multiple data centers, necessitating the development of vertically scalable cloud architectures ranging from entry-level CPUs to data center nodes. Consequently, a more decentralized solution is necessary, one that allows for data processing prior to transmission and storage. (Giannoutakis et al., 2020).

As a result, a massive volume of data must be transported throughout the network, stored, and handled efficiently by the receivers. The diversity of linked devices is tremendous and may be classified on various dimensions, including computational performance, storage capacity, network needs, communication protocols, and energy usage. This heterogeneity benefits a wide variety of IoT-connected device-based applications. Their needs grow in proportion to their numbers, making it increasingly difficult and significantly more demanding to satisfy the many demands that the computer system must be capable of answering. (Giannoutakis et al., 2020).

Decentralized Cloud Computing Model

The quantity of data generated grows exponentially with the number of connected devices. In the cloud, traditional centralized data centers are no longer an effective or sustainable option. It is critical to bridge the gap between resources and processing power at the edge. The definitions of key technologies in IoT described in Table 4 require to be broadened to the context of decentralized cloud computing model, which is why they are repeated.

Edge Computing

Edge computing is the process of transferring the intellect, processing power, and intercommunication capabilities of an edge gateway directly to the devices. It is often not associated with any cloud-based services, preferring to focus on the IoT device side. Appropriate resource management is critical, since task offloading might result in increased expenditures due to downtime and energy costs. Excessive resource processing on the servers may influence the performance of tasks in a system with a large number of users. Task scheduling, resource allocation and provisioning, and workload balancing are just a few of these metrics. (Mijuskovic et al., 2021).

Fog Computing

Fog computing is a methodical, highly virtualized, secured, and network-connected computing platform. It connects endpoints to conventional Cloud computing data centers through computing, storage, and networking services. It is also a novel way of digesting massive amounts of data and distributing it to customers through geographically spread platforms. Between the cloud and the end devices, fog computing makes use of nodes where intelligence may be found. (Mijuskovic et al., 2021).

These resources are provided by fog nodes, which are network nodes positioned near the edge. Fog nodes may be clustered vertically or horizontally. These intelligent nodes are assigned to serve as base stations or access points. By removing intelligence from the cloud, fog computing enables near proximity processing of IoT data. (Mijuskovic et al., 2021).

Additionally, broader applications such as distributed smart building control, Cyber-Physical Systems (CPSs), Wireless Sensor and Actuator Networks (WSAN), connected vehicles (CV), and IoT are also examples of larger applications. On the one hand, fog computing has the potential to significantly reduce network traffic by relocating data processing to the edge. On the other hand, security of data and edge cloud infrastructure at the edge may be a concern. Fog computing is not a replacement for Cloud computing. (Hajibaba and Gorgin, 2014; Pan and McElhannon, 2018; Mijuskovic et al., 2021).

Mist Computing

According to Iorga et al. (2018), mist computing is a simple and lightweight implementation of fog computing. It is also physically closer to the edge network and devices, which reduces end-user latency. Although its presence is not required, it is often implemented as a second fog computing layer, closer to or even on the same layer as end devices, according to Giannoutakis et al. (2020).

Serverless Computing

Serverless Computing is the process of developing, executing, and delivering applications and services entirely without concern of the server-side. Serverless does not indicate that no servers are used. FaaS and event-based programming are identical with serverless computing, which means an event may initiate the execution of a single function or a collection of functions simultaneously. (Giannoutakis et al., 2020).

MCC/MEC

The Internet-of-Things (IoT) applications are getting more complicated, needing a large amount of processing resources and a low latency requirement. IoT applications are getting more complex, requiring a high level of computing power and a low latency. Mobile cloud computing (MCC) is an efficient technique to overcome processing capacity constraints by shifting complicated activities from mobile devices to central clouds. Additionally, mobile-edge computing (MEC) is an emerging technology for lowering data transmission latency and saving energy by offering timely services. (Wu et al., 2020).

MEC makes use of computational resources in close proximity to IoT devices. MEC connects IoT devices to edge servers rather than to cloud servers directly. As the amount of communication and compute required for new city IoT applications increases, techniques based on deep learning may assist offload decision-making, dynamic resource allocation, and content caching. MEC and MCC offloading choices are getting more sophisticated as new technologies and paradigms emerge. MEC is seen as a vital technology for enabling the transition to a future world of IoT and 5G. Its prime objective is to allow cloud computing capabilities and an IT service environment at the edge of mobile-cellular networks. (Pan and McElhannon, 2018; Wu et al., 2020).

Challenges

As is the case with any paradigm, edge computing is not without its challenges. Thus, the story of edge computing is far from done. Indeed, it has not yet begun. When businesses adopt edge computing, one of the primary roadblocks they will confront is complexity.

Several examples, such as orchestration obstacles, how to avoid vendor lock-in while constructing edge networks, and the complexity of developing secure and compliant edge environments, are provided in the following sections.

According to a Redhat article (Red Hat, Inc., 2022), edge computing puts content, data, and processing closer to applications, objects, and people. It promotes the placement of workloads and capabilities in ways that optimize latency, bandwidth, autonomy, and security. Edge computing is not a substitute for cloud computing; instead, it will complement and complete it, as cited by Ruth (2020).

According to the Linux Foundation (2021), IT and OT are converging, with this trend being most apparent at the edge. Historically, OT solutions for managing and automating industrial equipment existed at the edge, while IT solutions were more centralized. Though these systems have been addressed independently, having an integrated IT/OT strategy would be advantageous. Nevertheless, there are concerns about who defines, controls, and maintains the edge device deployment to guarantee synergies with data and analytics and the broader digital business strategy. (Industrial Internet Consortium, 2018).

Giannoutakis et al., (2020) study identify five main barriers to loT-Cloud adoption:

- **Interoperability** - platform-based applications should be able to combine various Cloud-IoT platforms' services and infrastructure.
- **Security and privacy** - although many people utilize public clouds, sensitive private user data is at risk. Private data may be required to be retained closer to devices and consumers in certain circumstances to enable computation at the edge or fog layer. Additionally, the difficulty of developing suitable encryption-decryption processes and algorithms that can scale across distributed clouds while using less energy exists.
- **Portability** - each application and service must be capable of migrating efficiently across platforms and following the traces and pathways taken by users across the network.
- **Reliability** - enabling instantaneous object and application communication while maintaining a high level of accessibility and connectivity.
- **Virtualization** - enabling resource provisioning and access to heterogeneous resources like GPUs and FPGAs.

Cloud computing is undergoing fast development as a consequence of the IoT. The combination of IoT with cloud computing shifts cloud computing less centralized and more distributed. (Giannoutakis et al., 2020).

Orchestration from the Cloud to the Edge

Edge computing and fog computing are terms that refer to the processing of data near to its source. NIST defines it "*as a local computing at the network layer encompassing the smart end-devices and their users. It runs specific applications in a fixed logic location and provides a direct transmission service.*" (Iorga et al. 2018). It is expected to lower the quantity of data delivered to centralized cloud data centers, hence decreasing network congestion and benefiting analytics and knowledge-based applications. (Svorobej et al., 2020).

According to Nygren et al. (2010, as cited by Svorobej et al., 2020), many IT businesses and DevOps adopters depend on cloud-to-edge orchestration to accelerate service delivery, simplify optimization, and decrease costs. A cloud orchestrator manages, coordinates, and organizes distributed computer systems, services, and middleware in an automated fashion. Cloud companies such as Microsoft provide orchestration tools such as Azure Automation as part of their services.

Orchestration is complicated because of the size, heterogeneity, and diversity of resource types in virtualized environments, as well as the inherent uncertainty of the underlying cloud environment. Uncertainties include the demand for resource capacity, failures, user access patterns, and application lifecycle activities. (Svorobej et al., 2020).

Virtualization is primarily used to orchestrate and manage a cloud-to-thing architecture. Virtualization has evolved away from traditional VMs and toward lighter options like containers. Various application packages, like containers, as well as technologies like Kubernetes architectures and Docker containers, have been introduced and suggested for Cloud-to-Edge clustering. Nonetheless, a cloud edge computing architecture description and associated orchestration strategy are required. (Svorobej et al., 2020).

Industry-standard and suggested orchestration designs specify best practices for high-level system design. To coordinate a Cloud-to-Edge architecture, a variety of resource management tools are available.

Containers alleviate cloud PaaS problems by enabling the instant spawning of self-contained programs. Containers are frequently referred to as PaaS building blocks due to their capacity to be created on both virtual and physical infrastructures. Dockers, as defined by Turnbull (2014, as cited by Svorobej et al., 2020), are frameworks based around container engines.

Kubernetes is a widely used open-source container management framework for managing containers and their associated services. Kubernetes' architecture is composed of three main components: Master Components, Node Components, and Add-ons. (Kubernetes 2021a).

According to Kubernetes (2021a), The Master Components act as the cluster's control layer, making global choices regarding scheduling, backup, and node pod deployment to hardware nodes. On a hardware node, the Node Components establish and maintain a viable Kubernetes environment. These components are installed on each node allocated for container hosting in the data center, providing network proxy functionality, ensuring a healthy container state, and enabling container runtime capabilities. Add-ons are a supplemental component group that extends the functionality of the Master and Node

Components by adding DNS, a web-based user interface, and resource monitoring. (Kubernetes 2021a).

Developer challenges

According to Giannoutakis et al. (2020), the implementation of cloud has proven to be a difficulty for the network's different providers. Cloud designs, infrastructure, and deployment have been created to tackle this problem.

Monolithic architectures are ones in which an application is comprised entirely of a single piece of software or a single platform. It is straightforward to develop and deploy such an application. However, when the software needs to scale up, the actual obstacles multiply rapidly. Service-Oriented Architectures (SOA) are designed to coordinate the interaction of services. The difficulty of coordinating all services via a centralized component is a considerable drawback, especially for large and complex projects. Microservices alleviate the SOA methodology's shortcomings. By splitting programs into distinct independent services, this method subdivides them into more granular components. (Giannoutakis et al., 2020).

Application mobility

Ikink (2021) asserts that application mobility helps enterprises to have more control over their apps. IT teams can transition seamlessly between hypervisors, public clouds, and container-based systems without risking data loss or prolonged downtime. By separating apps from their runtime environment, IT teams may migrate them between different environments. Businesses will continue to prioritize application mobility since it enables the design of modular and portable apps.

Among the advantages of application mobility are the following:

- **Not reliant on a single platform**: Datacenter administrators may choose the optimal infrastructure for their needs and quickly relocate if circumstances change.
- **Businesses may adopt future technologies** such as containerization and public cloud services without fear of being locked onto an outmoded platform.

- **Concentrate on the application**: Administrators may select from a variety of basic platforms that do not need substantial training, enabling them to concentrate on the application's management. Ikink (2021).

According to a study by Section (2021), distributing workloads over a distributed edge brings a completely new set of challenges and issues. Managing hundreds or more deployment endpoints demands a distinct knowledge base and skill set.

Developers have been intimately acquainted with cloud deployment during the past cloud era. There are several reasons for the cloud's appeal, but the primary one is that it simplifies the administration of services for end-users by combining computation, storage, and delivery operations, according to the study.

The study determined that the developer experience is very same regardless of which cloud environment is used, AWS, Azure, or GCP. Cloud workloads often involve the following for a developer:

- Determining the locations with the largest user concentrations and picking a single cloud region that would deliver the most remarkable performance to the most significant number of users.
- Creating a connection to the codebase, which is stored in a code repository tooling.
- Using continuous integration (CI) and continuous delivery (CD) tooling to automate build and deployment.

When all code and microservices are contained under a single deployment destination, these operations are pretty trivial. What happens, though, when many more edge endpoints are added, each serving a separate microservice at a different time? How to determine which of the code's edge endpoints should be active at any given time? More significantly, how to handle continuous orchestration among these nodes when the infrastructure is diverse and comes from a variety of different providers? (Section, 2021).

Management of Code and Configuration

Each application is distinct. To meet the needs of their applications, developers want extensive, code-level control over edge settings. Simultaneously, they want simplified, efficient processes in order to sustain a high rate of innovation while still ensuring safe, trustworthy application delivery. (Section, 2021).

With several market participants fighting for a piece of the edge computing pie, there are numerous factors to consider for improving the developer experience to accommodate edge complexities. According to the study, "*many conventional CDNs, for example, have hard-coded proprietary software into their products*," leaving little room for flexibility. Thus, developers may find themselves cornered by outdated CDNs that provide edge services, forcing them to add additional solutions that ultimately negate some of the benefits they wanted with the CDN solution in the first place. Additionally, developers are increasingly relocating more application functionality to the edge in order to improve speed, security, and cost-efficiency. Section (2021).

Developer Workloads Are Shifting to the Edge

Numerous and diverse workloads are being evaluated for deployment on edge. According to the Section (2021) study, the following are a few instances of developer workloads shifting to the edge:

- Micro APIs - Host tiny, tailored APIs at the edge using GraphQL for use cases such as search or thorough content discovery, resulting in more rapid response times to user queries while reducing expenses.
- Headless Commerce - By separating the presentation layer from back-end eCommerce processes, it is possible to bring more services to the edge, resulting in a more customized user experience, improved performance, and operational efficiency.
- Placing the whole application at the edge - An increasing number of developers are experimenting with the notion of putting the entire program at the edge. Rather than resorting back to a single centralized origin, hosting databases alongside edge applications and then synchronizing across distributed endpoints is fast becoming a

reality that, as edge computing matures, it has all the abilities to become the norm of tomorrow.

Developers demand versatile solutions that facilitate code distribution across programming languages and frameworks to continue sophisticated workload migration to the edge. (Section, 2021).

Edge Developer challenges

According to the Section (2021) study, developers confront various obstacles while working with the edge, including the following:

With developers writing code in various runtime environments, there must be methods to facilitate code portability. Developers cannot be asked to rewrite their software in order to fit it inside a predefined framework. Rather than that, platforms and services for multi-cloud and edge must be adaptable to a variety of architectural styles, frameworks, and programming languages.

Developers must have a comprehensive view of their application's condition at any point in time. When considering distributed delivery nodes spread over a heterogeneous infrastructure set from several providers, observability gets exceptionally complicated.

A viable multi-cloud/edge observability solution should be capable of presenting data from several locations and infrastructure providers through a single pane of glass. This awareness is critical for developers to acquire insight into their apps' whole development and delivery lifecycles. By deploying a proper centralized telemetry system, engineers and operations teams may analyze performance, diagnose errors, monitor traffic trends, and share value and insights with relevant parties, according to the study.

Along with configuration flexibility, control, and robust observability tools, developers require the ability to manage their application lifecycle systems and processes effortlessly. This is relatively trivial with a single developer or small team in charge of a minor, centrally controlled codebase. However, when an app is divided into dozens of microservices that are controlled by many teams, and when the application architecture has a varied set of deployment methods, this may become significantly more difficult and impair the pace of development cycles. (Section, 2021).

Numerous teams simplify operations via the use of GitOps workflows. GitOps is a method for cloud-native apps to accomplish Continuous Deployment. It emphasizes on a developer-centric experience while managing infrastructure utilizing technologies known to developers, such as Git and Continuous Deployment tools. (Section, 2021).

While an organization's duties and control of various components of an application's code may be compartmentalized, the code must flow into a unified codebase. To enable developers to deploy more services to the edge, they want technology that is GitOps-based and supports an integrated edge-cloud application lifecycle. Developers also want flexibility and control when incorporating edge deployment methods into current CI/CD pipelines as an essential element of GitOps operations. (Section, 2021).

The Section study recommends that the key to increasing acceptance of edge computing is to make the programming experience at the edge as similar as possible to developers, drawing directly on cloud deployment themes.

The Future of Cloud Computing - A Distributed Cloud Model

As the cloud's role in business operations becomes more established, the ramifications of latency concerns become more severe. Gartner estimates that the distributed cloud may be able to address some of these difficulties. (Gartner, Inc., 2020a).

According to Gartner, the "*distributed cloud is the first cloud model that incorporates the physical location of cloud-delivered services as part of its definition. Historically, location has not been relevant to cloud computing definitions. In fact, location has been explicitly abstracted away from the service, which inspired the term cloud computing in the first place.*" (Gartner, Inc., 2020b).

Edge computing and public and hybrid clouds are the origins of the distributed cloud. Public cloud providers have offered numerous regions and zones for many years. Public cloud services may now be delivered to additional physical locations through bundled hybrid solutions, such as the edge. The distributed cloud architecture establishes strategically located cloud computing, storage, and networking substations that serve as shared cloud pseudo-availability zones (Gartner, Inc., 2020b).

According to Sahai (2021), by 2024, at least some distributed cloud services will be offered by the majority of cloud service platforms. However, in order for this to function, organizations must have a single perspective. The shared responsibility paradigm of the cloud provider must enable distributed cloud governance, which is currently not practicable.

Everything is getting more intelligent as a result of the IoT and wearable computing. Smart objects create a huge volume of data, which must be sent to the cloud to be processed. This centralized cloud approach introduces latency, making it inappropriate for time-sensitive and bandwidth-intensive applications. Using the distributed cloud, computing instead happens at the user's location, at the user's home, or inside the user's workplace. The user may now execute data-intensive and latency-sensitive AI, IoT, efficiency, and other applications at the edge. Businesses may harness the distributed cloud's capacity to build new apps and optimize current ones. (Sahai, 2021).

Sahai also states that the cloud is here to stay, and it will keep evolving to satisfy the ever-changing and increasing needs of applications and businesses. Distributed cloud, with a cloud-agnostic approach, is the ideal architecture for a future where smart devices are ubiquitous. (Sahai, 2021).

Gartner emphasizes future-proofing stating "*it is imperative that enterprises prioritize a distributed cloud-based solution as the default and future-proof edge solutions by relying on partnerships and ecosystems over a single-vendor approach*." (Gartner, Inc., 2021b).

2.5 Microsoft Azure in-depth case study

This chapter goes further into the Microsoft Azure platform. Providing an overview of the central cloud, edge, IoT, and security services needed on the ongoing data revolution, as well as a brief review of cloud sustainability.

2.5.1 Microsoft Azure in general

According to Copeland et al. (2015), cloud computing enables economies of scale, which may be leveraged to lower the cost of IT operations. Additionally, it helps organizations of all sizes to attain a high degree of availability and resilience. Customers may employ cloud computing to scale up and down as needed dynamically. Cloud computing includes all of the capabilities essential to increase the effectiveness of IT operations.

Azure, Microsoft's cloud computing platform, was designed to offer cloud computing services. However, it is not the outcome of an overabundance of computational capacity that might have been used for other uses. Microsoft Office 365, which is significantly reliant on the Azure infrastructure, was designed right into it. Microsoft has the requisite mix of technologies, infrastructure, and financial commitment necessary to bundle almost every facet of a SaaS, IaaS, or PaaS product. (Copeland et al., 2015).

Azure is designed to comply with industry-recognized information technology standards, especially cloud computing services. The Microsoft Trust Center maintains an extensive list of Azure platform certifications. Each service may have its own set of licenses and usage models since Azure is an assemblage of cloud computing services. (Copeland et al., 2015).

Azure IaaS Services

The infrastructure for Azure Virtual Machines (VMs) is designed to support both Microsoft and non-Microsoft technologies. Each Azure VM supports up to 480 CPU cores. Microsoft offers 12 hardware platforms for VM hosting, each with a distinct purpose. (Copeland et al., 2015; Soh et al., 2020).

Azure Storage is an umbrella term for a collection of Microsoft-managed offerings. Among these are, according to Soh et al. (2020), "*Azure Blobs, Azure Data Lake Storage, Azure Files, Azure Disks, Azure Archive, Azure Queues, and Azure Tables*," to name a few of the services available. Azure Storage can be scaled up or down in seconds while paying only for the resources used. (Soh et al., 2020).

At rest, all Azure Storage is encrypted with 256-bit AES, one of the strongest known encryption algorithms that comply with FIPS 140-2. By default, all Azure Storage accounts are encrypted regardless of the storage tier. Options for Azure Storage redundancy make use of Azure Storage Encryption. It is entirely free and has no adverse effect on storage performance. Additionally, no scripting or setting is required to use it. It may be configured to utilize either keys managed by Microsoft or keys managed by the customer via Azure Key Vault. (Soh et al., 2020).

Azure Storage delivers highly available, secure, durable, scalable, and redundant storage. Azure Blob storage may be used to store vast volumes of unstructured data and can be utilized to expose data publicly or privately. Azure Data Lake Storage Gen v2 enables the storage of vast volumes of data on the existing Azure Blob storage platform while being completely interoperable with a number of analytic systems. (Bansal, 2020).

Azure offers 19 networking services, including Azure Load Balancer and Azure Virtual Network, as well as networking services with unconstrained scalabilities, such as Azure's Virtual WAN. Almost any kind of connection, security, or availability strategy is configurable using Azure networking services. (Soh et al., 2020).

One may use one of three network connection methods to connect networks to Azure: ExpressRoute, site-to-site virtual private network (VPN), or point-to-site VPN. Numerous Azure services may be quickly made accessible through Azure Networks. Virtual networks (VNets) are analogous to physical networks. Azure VNets support both public IP addresses and private IP address ranges. (Copeland et al., 2015).

Additional "*Control it as if it were an IaaS, scale it as if it were a PaaS*" services worth bearing in mind include VM availability sets, VM scale sets, Azure Batch, and Azure Service Fabric. (Soh et al., 2020).

Azure CaaS Services

Azure Container Apps is Azure's fully managed Containers As a Service (CaaS) serverless computing offering. Azure Container Instances (ACI) is Microsoft Azure's most fundamental, straightforward, and simple-to-manage container solution. ACI enables rapid deployment of containerized applications without the need for additional configuration or provisioning of additional infrastructure, whereas the Azure Container Registry (ACR) service provides an enterprise-grade private container registry for storing and managing Docker container images. (Ifrah, 2020)

Microsoft Azure Kubernetes Service (AKS) is an enterprise-grade orchestration solution capable of scaling and managing massive deployments of containerized workloads. AKS's autoscaling capabilities enable it to rapidly expand the number of nodes, storage volumes, and networking without requiring manual or human interaction. AKS tools are tightly integrated with Azure DevOps and Visual Studio products, allowing the user to link the application straight to these tools and deploy it with ease. AKS is fully integrated with ACR and enables rapid container image retrieval. (Ifrah, 2020).

Azure PaaS Services

According to Soh et al. (2020), Azure PaaS offerings include services such as:

The Azure App Service platform is suitable for a wide range of software choices. The term "Azure Web Apps" refers to a managed application that is hosted on the Azure App Service. Azure Web Apps makes it possible to quickly deploy web-based apps as well as rapidly scale up or down in response to the load.

Azure DNS offers worldwide name resolution through Microsoft Azure's cloud-native infrastructure. Since the mid-1980s, DNS (Domain Name System) has been a critical Internet component.

Azure Content Delivery Network (CDN) delivers online services and content effectively with the smoothest possible user experience and response times. By using point-of-presence (POP) edge server repositories, the Azure CDN caches content and minimizes network latency.

Azure Private Link is a relatively new feature that enables connecting to PaaS services over an IaaS virtual network (VNet). This assigns a private IP address to the workload rather than a public IP address, enabling connection to the workload through Azure VPN tunnels or ExpressRoute private peering.

Microsoft Azure Database Services offers more than just Microsoft SQL Server. MySQL, MariaDB, and PostgreSQL are all popular commercial database hosting offerings. Additionally, Azure provides other services that aid with access and migration. Microsoft SQL services consist of two components: a "SQL managed" service and an Azure SQL Database service. In 2017 Microsoft announced the availability of Azure Cosmos DB, Microsoft's cloud-based proprietary database service that is globally distributed, multimodel, schema-agnostic, horizontally scalable, and has extremely low latency. Cosmos DB is a NoSQL database at scale. (Soh et al., 2020).

Azure FaaS Services

Microsoft's FaaS is called Microsoft Azure Function. It provides the developer with the benefits of event-driven serverless architecture, allowing them to create microservices using Azure applications and other third-party apps. (Microsoft, 2021g).

Azure SaaS Services

Microsoft's most popular SaaS offerings are all built and hosted on Azure and include products such as Microsoft Office 365 and Microsoft Dynamics 365.

Azure has many other services which might not fit well straight into only one as-a-service slots, but are worth mentioning, such as:

Azure Active Directory (AAD) is a Microsoft Azure service that enables cloud-based identity and access management. AAD serves as the basis for identity management in Microsoft Office 365. AAD has grown and evolved comprehensive identity-as-a-service solution. Users cannot access resources without the ability to authenticate. AAD delivers single sign-on (SSO) for over 2,400 pre-federated applications. Multi-factor authentication (MFA) in Azure adds an additional layer of protection. MFA may be configured to authenticate a user's identity via the use of a physical key, a mobile application, a text message, or a phone call. (Copeland et al., 2015).

Azure Service Bus operates as a communications foundation, connecting devices, apps, and services to other cloud-based applications and services and transferring data across them. Service Bus Queues operate on a one-to-one basis as the communication mechanism for Service Bus Topics is one-to-many. Each queue functions as a broker, storing delivered messages until the recipient collects them. Azure Service Bus allows enterprises to pursue a hybrid data center strategy that is both advanced and cost-effective. (Copeland et al., 2015; Bansal, 2020).

Azure Automation is a feature in Azure that enables automating and orchestrating Azure administration operations. Runbooks in Azure Automation may be created and tailored to conduct the system operations necessary for practical work. Automation runbooks can accept arguments from the user, generate output, and even invoke child runbooks. (Copeland et al., 2015).

Azure enables the management of custom, public, or proprietary application program interfaces. One method to expedite the usage of a developer platform is to offer an Azure-connected API management procedure. Azure API Management enables companies to grow by enabling the usage of their APIs by other organizations, customers, partners, and private developers. (Copeland et al., 2015).

Machine Learning is the computational effort that software does when it analyses sampled and historical data. Microsoft Azure Machine Learning enables the building of prediction models and analytical experiments. Python is an open-source programming language that enables the creation of solid applications, while R is a data scientist and statistician favorite. (Copeland et al., 2015).

Edge computing and multi-cloud from Azure perspective

According to Lee et al. (2021), edge computing merges the cloud's computational capacity with the capabilities of IoT devices. At edge locations near the data, edge computing may be used to operate virtual machines, containers, and data services to acquire real-time insights and decrease latency.

The widespread use of intelligent sensors and devices, along with cutting-edge cloud technologies like AI and machine learning, results in IoT devices that are very responsive to local changes and contextually aware. Additionally, due to the distributed nature of edge computing systems, there are security advantages since a single interruption is unlikely to jeopardize the whole network. (Lee et al., 2021).

A multi-cloud strategy entails using a variety of cloud computing services from several different cloud providers in order to get the optimal combination for a given activity or to leverage offers in certain areas, whether public or private clouds. For example, clients may adopt a multi-cloud approach in order to adhere to regulatory or data sovereignty requirements in a number of jurisdictions. (Lee et al., 2021).

Multicloud models may grow quite complicated due to the need to manage several platforms. Microsoft Azure offers tools to assist in the smooth operation of hybrid clouds, one of which is Azure Arc, a multi-cloud management platform. Azure Arc unifies administration and services across hybrid, multi-cloud, and edge environments with a single control plane, delivering a consistent state across resource environments and infrastructures. (Lee et al., 2021).

Azure Arc and Azure Stack

According to Lee et al. (2021), Azure Arc and Azure Stack are two components of Microsoft's solutions and technologies for hybrid and multi-cloud computing. Working together, they enable to:

- Modernize on-premises or edge apps with cloud-native technologies.
- Manage, regulate, and protect servers, Kubernetes clusters, and applications using a single control plane.
- Maintain Azure services on any infrastructure, providing automation, centralized administration, and security.
- Extend Azure computation, storage, and AI capabilities to the IoT and other edge devices to get real-time insights.

Azure Arc

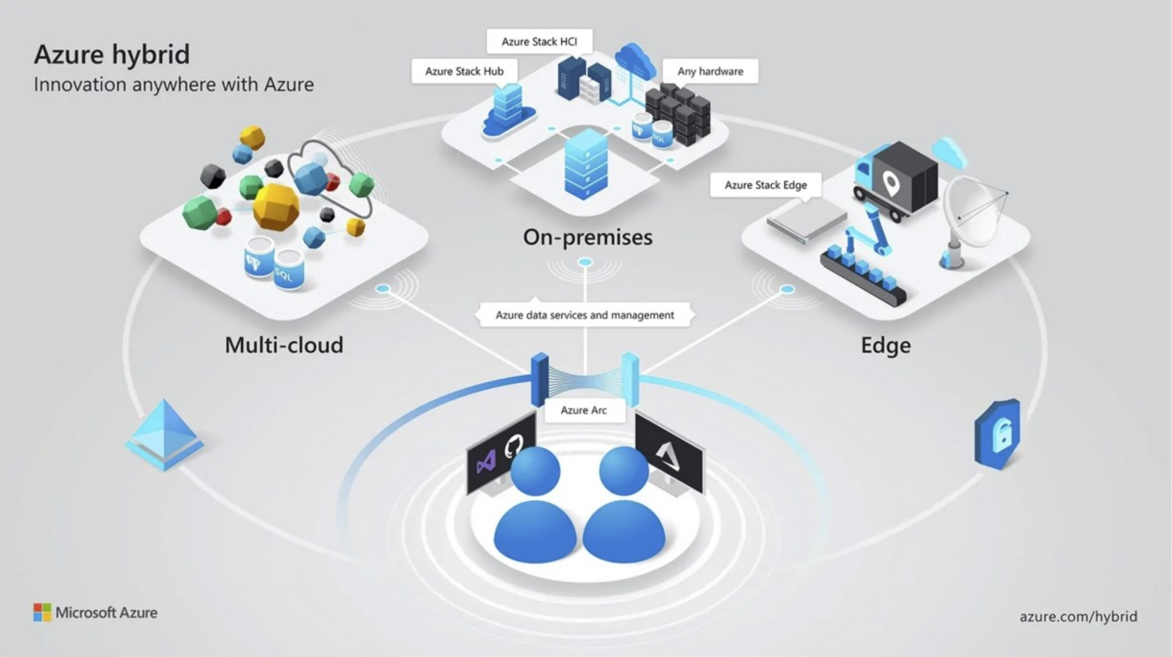


Figure 12: Azure Arc (Used with permission, Microsoft, 2021a).

Organizations must manage more complex information technology systems as use of cloud, non-cloud, on-premises, and off-premises technologies keeps growing at an exponential pace. Several clouds and ecosystems need the use of numerous dissimilar management tools and learning curves. Additionally, conventional technologies may be insufficient to enable cloud-native variants of operational concepts such as DevOps and ITOps. (Lee et al., 2021).

Azure Arc enables the deployment of Azure services and the extension of Azure management to any location. It is a centralized management platform for multi-cloud, on-premises, and edge resources. This enables the management of virtual machines, Kubernetes clusters, and databases in the same way they would be managed in Azure. (Lee et al., 2021).

According to Lee et al., (2021) and Microsoft (2021a), Azure Arc's features include things such as:

- Insight into operations and compliance via a single-pane-of-glass.
- The capability to construct and create hybrid applications at scale without jeopardizing central visibility and control.

- Confidently write and deploy apps from a central place to any Kubernetes deployment.
- With Azure Arc, one may continue and expand the use of IT-managed processes and services for business.
- Uniform server inventory, administration, governance, and security across environments are enabled by the core characteristics.
- By setting Azure VM extensions for Azure management services, the servers can be monitored, secured, and updated.
- Control and manage Kubernetes clusters at scale with Configuration as Code (CaC), which enables application and configuration deployment straight from source control using GitOps methods.
- The usage of Azure Policy to manage and configure the Kubernetes clusters in an automated manner.

Azure Arc addresses a number of the issues that businesses have when data is dispersed across hybrid cloud architecture. Azure data services supported by Azure Arc provide enterprises' data infrastructure with cloud flexibility. Customers may dynamically scale their databases in the same manner they can in Azure, depending on the available capacity of their infrastructure. This functionality may be used to address burst situations with volatile requirements, such as those that need real-time ingestion and querying of data at any size with a less than a second response time. (Microsoft, 2020a).

Azure Stack

Azure Stack is a solution family comprised of three products that enable the extension of Azure services and capabilities to any location. Additionally, hybrid applications may be created and operated reliably across location and environment boundaries, addressing the needs of various workloads. (Lee et al., 2021).

Microsoft Azure Stack Hub is a cloud-native integrated solution that enables the on-premises use of Azure cloud services. Stack Hub can be used in scenarios involving data sovereignty, regulation, compliance, and the modernization of applications. It is a private, independent cloud infrastructure that enables both entirely isolated and connected-hybrid cloud scenarios while retaining operational consistency with Azure. Using Azure Stack Hub, one may construct applications in an Azure environment for on-premises deployment without modifying any code. Data may be processed locally in the Stack Hub installation and then aggregated on Azure for further analytics at a suitable moment, eliminating latency difficulties or a persistent connection (Lee et al., 2021).

According to Lee et al. (2021), Azure Stack HCI is a hyper-converged infrastructure (HCI) that modernizes datacenters by combining contemporary software-defined storage and networking technologies in conjunction with Hyper-V for computing. Utilizing Azure Stack HCI to renew and maybe consolidate outdated virtualized hosts provides a number of potential advantages. It may improve scalability and make managing and securing the customer's environment easier. Additionally, by replacing traditional SAN storage, it helps minimize the footprint and total cost of ownership. (Lee et al., 2021).

Azure Stack Edge is a cloud-managed edge device that can be used to execute AI/ML, IoT, and edge computing applications. (Lee et al., 2021).

Infrastructure as a Code (IaC) in Azure

According to Lee et al. (2021), the primary advantage of IaC is that it automates the creation, maintenance, and setup of various resources and configurations. Automation is helpful at any step, and it eliminates the human aspect while providing some critical benefits such as:

- Scheduling automated deployments.
- Automated deployment smoke testing ability.
- The ability to construct a reproducible workflow.
- It is possible to create pure self-healing applications.
- The capability to revert modifications.
- The Resource Manager facilitates resource tagging.
- Azure Resource Manager (ARM) manages the resource dependencies.

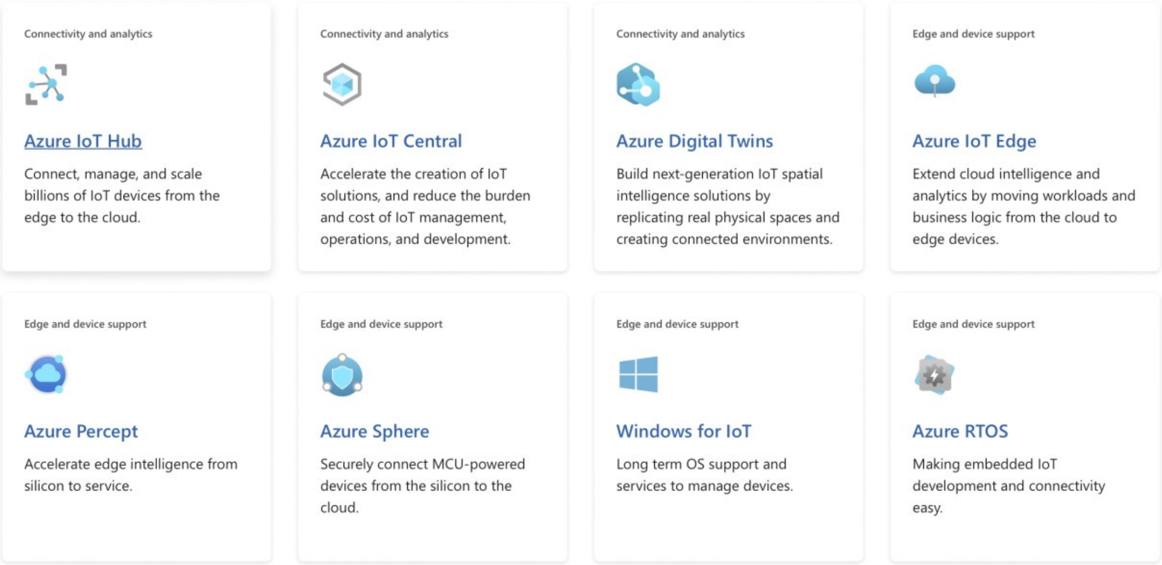
2.5.2 Microsoft Azure IoT and Edge services

Microsoft Azure IoT

Sensors convert the physical qualities of their environment to digital values in the IoT ecosystem. Each day, hundreds of times, this data is transported over established communication routes to the cloud or other suitable storage. Utilizing events or machine learning, these data are analyzed and used to create meaningful judgments. To make sure everything works, there has to be a scalable, secure, manageable, practical, reliable, and available architecture. The data saved should integrate easily with other internal or external components to be utilized efficiently. Another difficulty is that various sensors and devices use a variety of different communication protocols. The IoT ecosystem consists of numerous moving parts, and no particular technology or solution is capable of meeting all of the criteria. (Bansal, 2020).

Azure IoT Solution Accelerators

Pre-configured with a variety of IoT solutions, Azure IoT Solution Accelerators is an open-source Paas offering designed to aid in the development process. These solutions may be customized to suit unique requirements. (Bansal, 2020).



Connectivity and analytics
Azure IoT Hub
Connect, manage, and scale billions of IoT devices from the edge to the cloud.

Connectivity and analytics
Azure IoT Central
Accelerate the creation of IoT solutions, and reduce the burden and cost of IoT management, operations, and development.

Connectivity and analytics
Azure Digital Twins
Build next-generation IoT spatial intelligence solutions by replicating real physical spaces and creating connected environments.

Edge and device support
Azure IoT Edge
Extend cloud intelligence and analytics by moving workloads and business logic from the cloud to edge devices.

Edge and device support
Azure Percept
Accelerate edge intelligence from silicon to service.

Edge and device support
Azure Sphere
Securely connect MCU-powered devices from the silicon to the cloud.

Edge and device support
Windows for IoT
Long term OS support and services to manage devices.

Edge and device support
Azure RTOS
Making embedded IoT development and connectivity easy.

Figure 13: Azure IoT services (Used with permission, Microsoft, 2022b).

Azure IoT Central

IoT Central is a SaaS offering that contains straightforward, preset templates to help customers get started fast with IoT projects. Numerous IoT Central application templates are available for vertical sectors such as retail, health, energy, and government. (Bansal, 2020).

IoT and IoT Edge devices are supported by all IoT Central applications; moreover, IoT Central offers preconfigured IoT Plug-and-Play devices from approved vendors as device templates. Capability models for devices are included in a device template, which defines the device's supported telemetry, properties, instructions, and cloud features such as the last time the device was serviced, custom device characteristics, and graphical user interface dashboards for monitoring and managing the devices. (Bansal, 2020).

Azure IoT Edge

IoT Edge does not fall under either the PaaS or SaaS category. According to Bansal (2020), "*It technically lets one run cloud intelligence, data analysis, and custom logic execution directly on IoT devices; when combined with edge processing, this can offer faster processing with low latency and reduce the time and cost of data travel to the cloud servers*."

Liu et al. (2019) stated, "*Azure IoT Edge attempts to bring cloud analytics to edge devices. These edge devices may be routers, gateways, or other computing-related equipment.*" Azure IoT Edge uses similar programming style as other Azure IoT cloud services, allowing users to transfer existing apps from Azure to edge devices in order to achieve lower latency. Additionally, Azure services like Azure stream analytics, Azure machine learning, and Azure functions may all be utilized on edge devices to execute complicated activities including machine learning, image recognition, and other tasks related to artificial intelligence. (Liu et al., 2019).

According to Liu et al. (2019), "*IoT Edge modules are containerized instances that execute client code or Azure services. The IoT Edge runtime manages these modules*." IoT Edge modules are where specific applications are executed. A module image is a Docker container image that includes the application's code. In the context of a Docker container, a module instance is a unit of computing that executes the module image. Due to the same programming style, these modules may execute the same Azure services or customized applications as in the cloud, providing the edge devices have sufficient resources. Additionally, since Azure IoT Edge is scalable, these modules may be dynamically deployed. (Liu et al., 2019).

On edge devices, the IoT Edge runtime acts as a manager. It is divided into two components: the IoT Edge Hub and the IoT Edge agent. The IoT Edge Hub serves as a local proxy for the IoT Hub; a cloud-based managed service and central communications hub. The IoT Edge Hub acts as a message broker, connecting modules and transmitting data to the IoT Hub. The IoT Edge agent is used to configure and monitor the modules of the IoT Edge. It obtains information about module deployment from IoT Hub, instantiates them, and checks that they are working correctly. Additionally, it communicates the modules' status to the IoT hub. (Liu et al., 2019).

Device administration is accomplished using a cloud-based interface called IoT Edge. Users may use this interface to design edge apps, transmit them to the device, and lastly monitor the device's status. (Liu et al., 2019).

A straightforward deployment technique for apps is choosing an Azure service or creating own code, compiling it as an IoT Edge module image, and deploying it to the edge device using the IoT Edge interface. The IoT Edge fetches the module image and creates the module instance after receiving the deployment information. (Liu et al., 2019).

The Azure IoT Edge is applicable to a wide variety of scenarios. It includes case studies in intelligent manufacturing, irrigation systems, and drone management. "*While Azure IoT Edge is open source, Azure services such as Azure functions, Azure machine learning, and Azure stream are billed*", reminds Liu et al. (2019).

Azure IoT Hub

Azure IoT Hub is a fully managed communications solution for IoT. It is bidirectional, allowing for command and control of communications between the cloud and the devices. It enables highly secure, scalable, and trustworthy communication between IoT devices and solutions. (Klein, 2017; Bansal, 2020).

The Azure IoT Hub is a set of Microsoft-managed cloud services for connecting, monitoring, and controlling IoT devices. SDKs for developing, managing, and monitoring IoT Hub devices are available in several programming languages. These cloud services are further subdivided into SaaS and PaaS offerings, each with its own set of trade-offs regarding customization and speed to market. HTTPS, AMQP, and MQTT protocols are supported by Microsoft Azure Internet of Things. (Bansal, 2020).

According to Klein (2017), the following advantages highlight the Azure IoT hub:

- **Scalability** refers to the capacity to accommodate millions of linked devices concurrently as well as millions of events per second. IoT Hub automatically scales in response to device additions.
- **Per-Device Authentication and Security**: Complete, fine-grained control over which devices are allowed to access the IoT solution and guarantee that cloud-to-device instructions are transmitted to the proper device.
- **Monitoring Devices**: Troubleshoot device connection problems with full operation logs. These logs provide information the management of device identities and device connectivity events.
- **Extensibility**: IoT Hub's extensibility allows for the addition of support for custom protocols.
- **Support for a variety of languages and platforms** through IoT device SDKs and device libraries. (Klein, 2017).

Additionally, the IoT Hub augments the architecture with critical functionality. It manages devices and device twins, as well as identify and authenticate users, upload files from devices, provision devices, and send cloud-to-device messages. To authenticate users, "*SAS tokens, individual X.509 certificates, or an X.509 Cert Authority are utilized*," according to Stackowiak (2019).

Once device credentials are authenticated, and messages begin flowing into the IoT Hub instance, messages may be forwarded to other Azure services, such as Azure Stream Analytics, Event Hub, Service Bus Queues and Topics, and Azure Storage. This capability is identical regardless of whether the device is an IoT Edge device or a standard IoT device. (Jensen, 2019; Stackowiak, 2019).

IoT Hub offers two subscription tiers: Basic and Standard. Standard has one benefit over Basic in that it supports cloud-to-device communication, streaming, and IoT Edge. The Basic tier is an excellent value if these features are not required. It is upgradeable at any moment without affecting service. There is a flat price per device. This is advantageous since there are no hidden costs, and the IoT project's cost is known in advance. Additionally, it is clearly calculable. On the downside, the price must be paid whether utilized or not. (Bansal, 2020).

Tier	**Edition Type**	**Price per Azure IoT Hub unit (per month)**	**Total number of messages/day per IoT Hub unit**
Basic	B1	$10	400.000
Basic	B2	$50	6.000.000
Basic	B3	$500	300.000.000
Standard	Free	$0	8.000
Standard	S1	$25	400.000
Standard	S2	$250	6.000.000
Standard	S3	$2500	300.000.000

Table 6: Azure IoT Hub Pricing West Europe Region 01/2022 (Microsoft, 2022c).

Azure Digital Twins

According to Stackowiak (2019), a digital twin is a technology that may be used to simulate the physical state of a device. Azure Digital Twins are constructed on top of the Azure IoT Hub platform.

Spatial intelligence graphs are used to construct an image of the actual environment. The schema may be used to represent connections among individuals, locations, and gadgets. Tenants, consumers, regions, building names or addresses, levels, spaces within floors, and gadgets are all possible representations of a building. Within these contexts, the data can then be queried (e.g., by location). (Stackowiak, 2019).

A digital twin, for example, may be used to interpret sensor data indicating the ambient conditions of a manufacturing facility. The telemetry data would be validated, matched, computed, and sent using the Azure Digital Twin. The computation is carried out within the confines of user-defined functions. Following that, the spatial intelligence graph may be used to perform location-based queries on data delivered to the Azure Digital Twin. (Stackowiak, 2019).

Through the IoT Hub APIs, Digital Twins allow backend service interactions. The information included in a Digital Twin document may be searched through those APIs, allowing situations such as device reporting on dashboards or monitoring of long-running operations across several devices. This is not straightforward to implement since each device requires a real-time request. With Digital Twins, queries may be performed in milliseconds, knowing that the return set represents the device's last reported status, not a real-time update. The majority of firms are ready to tolerate some degree of delay. (Jensen, 2019).

Backend services may communicate with the Digital Twin using the IoT Hub's device API. Services may read and update device tags, read and update desired properties, and read and query reported properties using this API. Additionally, the device may read desired attributes as well as read and update reported properties. (Jensen, 2019).

Azure Stream Analytics

According to Bansal (2020), "*Azure Stream Analytics is a fully managed serverless (PaaS) real-time data analytics and event processing engine*." It is intended for use with streaming data from a variety of sources, including Azure Blob storage, IoT hubs, and event hubs. It may be implemented in the cloud or containers on edge devices. Azure Stream Analytics' primary function is to correlate data from numerous sources, filter out irrelevant or poor data, transmit data, aggregate the data, and lastly, analyze the data to provide useful conclusions for additional actions or triggers. All transformations are done in a language similar to SQL. "*The final result is call output, and Azure Stream Analytics can output results with ultra-low latencies to various Azure services; it can store output in SQL Database and Cosmos DB, can store output in Azure Storage and Azure Data Lake Gen1 Storage, can feed output for further processing to Service Bus Topic/Queue and Azure functions, or simply create a data source for a real-time dashboard in Power BI,"* according to Bansal (2020).

Azure Time Series Insights

One of the primary objectives of any IoT project is to transform data into meaningful insights, such as enhancing operational efficiency, identifying irregularities, and spotting hidden patterns for rapid decision making. Real-time data insights provided by the Azure Time Series Insights (TSI) service assist in doing this.TSI is a fully managed product explicitly designed for data produced by IoT devices. It ensures that data arrives in a consecutive sequence. TSI refers to them as input event sources, and the two accessible alternatives are Event Hub and IoT Hub. Trends and data are automatically recorded, combined, and graphically visualized by TSI. It enables on-demand querying and viewing using the TSI explorer. (Bansal, 2020).

Azure Databricks

According to Stackowiak (2019), "*Azure Databricks is an analytics platform built on Apache Spark that enables the preparation of in-memory data and the training of machine learning models.*" Raw streaming real-time data from the IoT Hub may be directly supplied into the Databricks cluster in an IoT solution footprint. Typically, data is retained in a data lake for the long term. Additionally, data may be pulled from permanent storage sources such as Azure SQL Data Warehouse, Cosmos DB, Azure Data Lake Storage, or other non-Azure data stores. (Stackowiak, 2019).

Databricks' collaborative workspace offers data exploration, notebook-based programming development, data visualization, as well as dynamic report generation. Python, R, Scala, and SQL are just a few of the programming languages that Databricks supports. Cluster autoscaling guarantees that a sufficient number of workers is always available to perform operations, even when a predefined number of workers is specified. When jobs are executed, Databricks will add more workers throughout these stages if particular sections of the pipeline are more computationally intensive than others and will remove them when they are no longer required. (Stackowiak, 2019).

Azure Data Lake Storage

The features of Azure Data Lake Storage Gen2 are derived from two prior storage services: Azure Blob Storage and Azure Data Lake Storage Gen1. According to Stackowiak (2019), "*Azure Blob Storage is a general-purpose object storage solution that is well-known for its cost-effective tiered storage*." Azure Data Lake Storage Gen1 enhanced Azure Data Lake Storage with file system semantics, directory security, and file-level security. (Stackowiak, 2019).

The cost-effectiveness of Azure Data Lake Storage Gen2 is achieved by integrating these features "*into a namespace that organizes files into a hierarchy of folders comprising underlying objects*," according to Stackowiak (2019). On directories and files, the Portable

Operating System Interface (POSIX) permissions may be specified. Additionally, security extensions such as access control lists (ACLs) and others are supported. Access to data is now more efficient than it was in the preceding generation because the Azure Blob File System (ABFS) driver was created with analytics in mind. (Stackowiak, 2019).

Azure HDInsight

Microsoft's cloud-based PaaS Hadoop platform, Azure HDInsight, was created in partnership with Hortonworks. It is presently most often deployed on Azure Data Lake Storage Gen2. (Stackowiak, 2019).

Cosmos DB

Cosmos DB, a globally distributed NoSQL database engine, has grown in popularity as a replacement for Azure Data Lake Storage settings. SQL, MongoDB, Cassandra, Azure Table Storage, and Gremlin are all supported APIs in Cosmos DB. Spark is supported for processing in-memory data stored in Cosmos DB. (Stackowiak, 2019).

According to Stackowiak (2019), "*the throughput and storage of Cosmos DB may be expanded elastically and independently across any number of Azure regions. Transparent multi-master replication guarantees 99.999 percent availability while also allowing for regional failover. The datastore is schema-agnostic. In addition, Cosmos DB automatically indexes all data.*" At the 99th percentile, read and indexed write latencies are assured to be less than ten milliseconds. All data, both at rest and in transit, is encrypted, and row-level security is provided. (Stackowiak, 2019).

Azure Synapse Analytics

Azure Synapse Analytics is composed of various components, one of which is the Synapse workspace. Additionally, the Azure Synapse Analytics workspace is composed of numerous components. When an Azure Synapse Analytics workspace is provisioned, additional components such as an Azure Data Lake Storage Gen2 account and a file system are created. Additionally, it will construct a Synapse SQL Pool that is Serverless or On-Demand. Microsoft integrated Synapse workspaces to Azure Synapse Analytics to assist in supplying and managing all of these components. The workspace serves as the hub for all interactions with Azure Synapse Analytics. (Shiyal, 2021).

When it comes to the computational engine, Azure Synapse Analytics offers three distinct alternatives. Synapse SQL is available in two flavors: dedicated and serverless. The third alternative is Synapse Spark, a Microsoft-developed version of Apache Spark optimized for Azure Synapse Analytics. Spark is the de facto compute engine of choice for the majority of Big Data analytics applications. Synapse Spark is one of the most intriguing new capabilities in Azure Synapse Analytics as a result of this. (Shiyal, 2021).

Azure Data Factory

Azure Data Factory is a cloud-native data orchestration solution for the Azure cloud. It has many native connectors, enabling it to connect to various data sources readily. Additionally, it is fully integrated with Synapse Studio, which means all the work can be done in the Synapse interface in order to design or maintain data pipelines. (Shiyal, 2021).

Logic and Function Apps

According to Bansal (2020), the Logic and Function apps are both examples of serverless computing. The term "serverless" refers to there is no need to maintain any infrastructure. Their interaction with IoT Hub is the same, even though both Azure platform services are distinct and valuable for different reasons.

The logic application is a collection of actions that form a logic process. To the untrained eye, it resembles a Microsoft Visio data flow diagram with groups of boxes linked by actions. The logic application is invoked through an HTTP request and has a single entry point. It may be scheduled to run at predetermined intervals, much like a service. Scalability is another feature of the Logic program. When the number of requests exceeds a predefined threshold, the number of logic app instances rises. It has a run history that allows for an in-depth examination of specific runs. (Bansal, 2020).

A function application involves developer abilities, and the code is deployed in the form of a modest collection of relevant and reusable methods. Azure functions are single-responsibility microservices that can be readily shared among services. Any language from .Net Core, Node.js, Python, Java, or PowerShell commands is supported. The execution history can be seen from the monitor tab. As general guidance, assess if using the Logic app satisfies the requirements; however, if the need is to build code with a significant level of flexibility, use Azure Function. Both Logic App and Azure Function are billed on a consumption-based approach and need no upfront investment. (Bansal, 2020).

Power BI

Microsoft's Power BI business analytics solution organizes, analyzes, and visualizes data. Collaborative work on dashboards, reports and data is possible. PowerBI includes charts, graphs, and many forms of business visualizations that professionals need daily. Users may generate personalized visualizations by utilizing ArcGIS maps and R and Python scripts. PowerBI may be linked to IoT Hub through Azure. (Bansal, 2020).

Azure Machine Learning and Cognitive Services

According to Bansal (2020), In the case of IoT Edge devices or IoT devices that create semi-structured data such as audio and pictures, AI models may be used to analyze the data. Azure Machine Learning is a cloud-based predictive analytics platform that enables machine learning models' training, deployment, and management. Similarly, with Custom Vision Service, custom image classifiers may be constructed and deployed. Both of these services are applicable to a vast array of industrial applications, including anomaly detection and predictive maintenance. Visual inspection of mechanical components, tool identification on a production floor, and part detection on a production line are all possible uses for Custom Vision. Likewise, machine learning can identify irregularities and recommend predictive maintenance for an engine based on auditory inputs. IoT Edge enables devices to handle data in real-time and avoids security breaches by bringing computing capabilities to the edge. (Bansal, 2020).

IoT and Bots

Bots may operate in conjunction with IoT devices to improve the sophistication of our lives. Microsoft Azure offers services for developing intelligent IoT solutions, which could be further extended by using Azure Machine Learning, Stream Analytics, and Power BI. Innovative applications may be constructed using the Azure Bots framework and IoT hubs. (Machiraju and Modi, 2018).

2.5.3 Sustainability in Microsoft Azure platform

As cited by Microsoft (2021d), the World Meteorological Organization reports that 2020 was the third-warmest year on record, with global temperatures increasing by an average of 1.2°C over preindustrial levels. Climate change and rising temperatures increase the frequency and severity of natural catastrophes and severe weather.

According to Microsoft (2021d), governments have acknowledged the critical need to cut greenhouse gas emissions significantly to safeguard people and the environment. Governments can use the potential of digital technology and data to accelerate their paths toward sustainability.

Additionally, digitalization may aid in the development of more sustainable IT and procurement choices. This is critical since the information technology sector is predicted to use 20% of global power by 2025. As a first step, governments should understand the effect of information technology acquisitions and may do so by using carbon footprint calculators such as Microsoft's Sustainability Calculator. Additionally, digital solutions may help governments monitor green procurement choices more effectively by leveraging the value of data and obtaining real-time insight into their supplier chains. (Microsoft, 2021d).

According to Lacy et al. (2020), migrations to the public cloud might result in cost reductions of up to 30%-40% on the total cost of ownership (TCO). More energy-efficient infrastructure, higher server utilization rates, and increased workload flexibility all contribute to the cost-effectiveness of public clouds. Additionally, migrations to the public cloud have the potential to offset 59 million tons of CO2 emissions each year. Furthermore, cloud migrations open up new options, such as sustainable energy transitions supported by cloud-based geographical analytics. This essential cloud trend will only intensify in the coming years.

According to Microsoft (2020b), during the last few years, a rising number of businesses have realized the advantages of cloud computing. Sustainability, on the other hand, must now be included in business initiatives. Due to cloud computing, businesses will be able to provide new services such as carbon reduction and responsible innovation. The growing demand for sustainability among businesses has prompted businesses to investigate cloud computing as a possible solution. The cloud has been highlighted as a critical component in achieving a net-zero future. By 2025, Microsoft plans to transition to a 100 percent renewable energy source, and by 2050, Microsoft plans to have entirely eliminated its historical carbon footprint.

Microsoft commissioned research to evaluate cloud-based apps' energy usage and carbon emissions with their on-premises equivalents. Microsoft chose cloud-based Microsoft Azure Compute, Microsoft Azure Storage, Microsoft Exchange Online, and Microsoft SharePoint Online because they together utilize around half of the energy in Microsoft datacenters. The study showed that the Microsoft Cloud consumes between 22 and 93% less energy than traditional enterprise data centers. When renewable energy purchases were taken into account, the Microsoft Cloud is between 72 and 98 percent more carbon efficient. (Microsoft, 2020b).

Gartner estimates, "*Hyperscalers are investing heavily in sustainable cloud operations and delivery, with the goal of ultimately reaching net-zero emissions within a decade, if not sooner. Gartner anticipates a growth in the availability of tools that aid businesses in calculating and reducing carbon emissions via efficient use of cloud services, similar to the tools that assist organizations now in optimizing cloud expenditure.*" (Gartner, Inc., 2022).

Additionally, according to Gartner, the top ten cloud providers (in terms of revenue) accounted for 70% of total IT expenditure on cloud infrastructure, platform, and application services. Cloud sustainability initiatives will begin with the industry's top cloud providers, which operate some of the world's largest data centers and are crucial for lowering IT-related carbon emissions. (Gartner, Inc., 2022).

2.6 Threat Landscape of Cloud / Edge / IoT

Cloud

As our reliance on the cloud increases, our susceptibility to cloud disruptions increases proportionately. Examples include military actions, natural catastrophes, and cyberattacks. Concerns are mounting that cyberattacks may soon become significant weapons of organized crime and tools of national strategy. (Satyanarayanan, 2017).

Cybercrime increased by 630 percent between January and April 2020, according to the McAfee report (2020, as cited by Ikink, 2021). Companies are learning that taking shortcuts on the cloud might make their organizational processes unclear, opening multiple access points for cybercriminals. The survey provided fascinating insights into the businesses that participated, including the following: "*Sixty-five percent of senior IT executives feel the most significant impediments are security and compliance risks, and twenty-eight percent of organizations prioritize security when selecting a cloud provider.*"

In another recent survey, conducted by Divvycloud (2020), nearly 2,000 information technology experts illustrate how organizations are embracing public cloud, multi-cloud, containerization, and other technologies. Too often, organizations make a false choice: either quickly and recklessly embrace the cloud or approach it cautiously.

While many businesses are making progress, many still struggle to achieve a complete security and compliance posture. Any organization can function efficiently and successfully in the cloud with appropriate planning and the right people, procedures, and technologies, the survey states.

Additionally, organizations do not have cloud security plans in place at the time of cloud adoption. The traditional role of security as a central controlling function has shifted—it is now everyone's duty to guarantee the safe deployment and management of cloud services.

According to the survey findings, just 7% of firms polled do not use any public cloud services, despite the fact that the great majority of enterprises have already embraced the cloud. Only 5% of firms have no intentions to utilize public cloud computing. Forty-two percent of information technology professionals are unaware of the frameworks their organization employs to comply with regulations and standards.SOC 2, CIS, FedRAMP CCM, NIST CSF, HIPAA, PCI DSS, and GDPR are just a few of the frameworks.

Cloud misconfiguration-related data breaches continue to be widespread, costing organizations an estimated $5 trillion in 2018 and 2019 alone, according to survey.

Fifty-nine percent of respondents who indicated their company had had a cloud services data breach or other security event in the recent 12 months said it resulted from a misconfiguration. According to the survey, "*almost half (49%) of respondents whose organizations use the public cloud said their developers and engineers at times ignore or circumvent cloud security and compliance policies*." Developers often feel constrained by security and compliance standards when it comes to developing and deploying new services.

Enterprises may deploy automated solutions after their security and remediation policies are prepared correctly. These controls enable businesses to identify mistakes that frequently occur when cloud security and compliance regulations are disregarded. Continuous security may be ensured by enterprises by implementing the essential people, procedures, and technologies concurrently with cloud adoption. (Divvycloud, 2020).

Edge

According to a recent Eclipse Foundation survey, the primary worry for IoT & Edge Deployments is data security. The findings show that 36% regard data security as a significant design priority when adopting IoT and edge technologies. The survey recommends that businesses deploy data security and sovereignty solutions across devices and apps, paying particular emphasis to their ability to maintain control over data flow and storage, for example, for data collected from IoT sensors and devices. (Eclipse Foundation, 2021).

According to recent research conducted by Hypothesis Group and Microsoft, safeguarding information technology infrastructure and assets would be a priority in 2021. Almost a third are worried about the security risks associated with IoT, notably with data privacy and network-level security. The principal objective is to prevent and identify data breaches to guarantee IoT projects' security, albeit no unique best practice is generally embraced. (Hypothesis Group and Microsoft, 2021).

According to Industrial Internet Consortium (2018), edge computing often expresses concerns about security. As with any other IT system, edge computing workloads and systems will be targeted by attackers. Physical insecurity adds complexity and a high number of attack surfaces. Innovations are necessary to monitor, manage, and safeguard globally distributed systems and limit unavoidable breaches.

According to Industrial Internet Consortium (2018), "*in edge computing implementations:*

- *Security must be built-in to each device and at every level of the architecture.*
- *Computing and networking endpoints must be monitored and managed.*
- *Latest patches must be applied.*
- *Attacks must be isolated and quarantined.*
- *Affected components must be able to be healed."*

Moreover, according to the Linux Foundation (2021), edge solutions will benefit significantly from the network- and SaaS-based orchestration systems in terms of manageability and security.

2.6.1 Different levels in security

Cloud

According to Sahai (2020), cloud security is built on a shared responsibility approach, which is often misunderstood. Cloud providers carry some accountability in this scenario. The remainder is entirely up to the consumer. The shared responsibility approach is perplexing since it varies according to the kind of infrastructure and cloud services. Cloud's three V's - volume, variety, and velocity - exacerbate this difficulty by drastically extending the assault surface.

Meanwhile, cloud velocity is rising as a result of the technologies that are displacing conventional infrastructure. Virtual machines will look archaic in a few years as containers, and serverless functions conquer the market. This is a very dynamic environment to manage and it is the consumer's responsibility to stay current with these developments and ensure that the security and compliance problems they raise are addressed. (Sahai, 2020).

Assessment is the first step in establishing a closed-loop corporate security strategy, according to Sahai. Cybersecurity risks are continuously developing, necessitating a regular assessment of the security environment. The second and third are monitoring and remediation. Monitoring entails flagging problems and raising alerts using security monitoring solutions such as security information and event management (SIEM) and cloud-native log management systems with business logic. Remediation comprises comprehending and using risk assessments in order to prioritize and resolve alarm-triggering situations. (Sahai, 2020).

In a recent security design guide by Cisco (2021), the top goals while creating secure public cloud solutions are as follows:

- **Visibility:** Ensure total view of users, devices, networks, applications, workloads, and processes.
- **Segmentation:** Limit the attack surface of attackers by preventing attackers from moving laterally.
- **Threat Protection**: Protect against breaches by strategically installing multi-layered threat sensors to identify, prevent, and react to attacks rapidly.

According to Cisco, these features operate in concert to safeguard cloud applications through many levels of security.

Any successful information security and compliance program should include risk assessment. This begins with determining what is critical and what is not. One must determine which assets are vital in order to safeguard them from dangers. Because everything cannot be resolved immediately, one must first prioritize the issues and then address them. In addition, micro-segmentation enables organizations to decrease their attack surface, safeguard essential applications, and strengthen their regulatory compliance posture. A well-defined micro-segmentation approach enables the uniform application of security controls across data centers and cloud platforms, according to Sahai (2020).

The cloud allows new and innovative technologies, and cloud computing is constructed on trust. Customers will avoid technologies and technology suppliers in which they lack confidence. According to Shinder (2019), "*Trust in cloud computing requires the ability to rely on services and data being available when they are required*." Reliability and resilience are essential components of the relationship of trust that exists between cloud providers and their clients. Additionally, trust is a vital aspect in the adoption and decision-making processes for cloud computing. There must be assurances that data will be available to those who need it to do their jobs and that it will be protected against illegal access, manipulation, or loss. As a result, data protection is a primary priority. (Shinder, 2019).

Edge

As previously mentioned, edge computing is the processing of data at or near its source rather than its transfer to the fog or cloud. For example, as stated by Shi et al. (2016), "*a smartphone serves as the interface between bodily objects and the cloud, while a gateway in a smart home serves as the interface between house objects and the cloud.*"

According to Ansari et al. (2020), while edge computing has several benefits and is employed in a range of contexts, it is not without security risks and issues. In the case of edge computing, the following aspects contribute to the expansion of the attack surface:

Hardware Constraints: Because the majority of edge computing gear lack the processing and storage capacity of a cloud or fog server, they are unable to operate specialized protection systems such as firewalls, making them more susceptible to cyberattacks.

Software Heterogeneity: Because the majority of edge computing gear communicates through a range of unstandardized protocols and operating systems, the process of creating a universal defense mechanism is complicated.

According to Xiao et al. (2019), the majority of these vulnerabilities are amplified by design flaws, implementation issues, and misconfigured edge devices and servers. Additionally, the majority of cyberattacks against edge computing infrastructure fall into one of four categories presented in Figure 14.

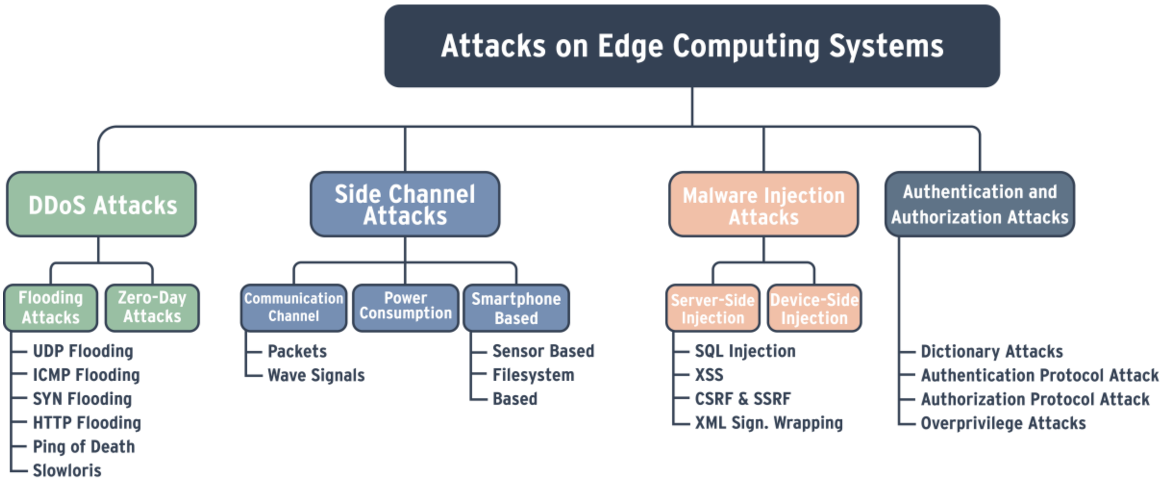


Figure 14: Different types of attacks against edge computing systems (Adapted, with permission, from Xiao et al., 2019 © IEEE).

Narayanan et al. (2020) describe these four main categories in more detail:

DDoS attacks are cyberattacks that try to disrupt ordinary services by flooding internet traffic. Flooding and logical attacks are two variants of this assault. A flood attack delivers an excessive amount of malicious packets to edge devices. Logical attacks involve sending malicious packets that mislead the target's application by indicating that all resources are occupied. DDoS attackers often target edge devices to be used as a weapon against something else.

In a **side-channel attack**, the attacker exploits publicly available data and correlates it with the user’s private data "secretly" to infer confidential data. Side-channel attacks may occur at any network node and use a variety of approaches. Intense research has been done on the vulnerability of DL/ML-based systems and devices to adversarial assaults.

In **malware injection attacks**, the attacker tries to access the victim's service requests and implant malware into the network or computer systems. These cyberattacks threaten data integrity and system security; specifically edge servers and devices are prone to this kind of attack.

Authentication confirms or verifies the entity's identity requesting services, whereas authorization ensures the rights and access of an entity within specific limits and bounds. **Authentication and authorization attacks** try to get access to resources using forged credentials.

According to Bhamare et al. (2020), "*the recent popularity of employing big data analytics and cloud computing for Industrial Control Systems (ICS), their security is still an open issue.*" Security breaches in real-time industrial platforms may be expensive due to the lack of adequate protection in unique multi-cloud systems. Recent years have seen an increase in the sophistication of malware designed to attack control systems, making it more challenging to prevent and detect assaults at the ICS component level. As a result, new intrusion detection approaches for ICS systems are required. Machine learning technologies have been shown to be successful in this case. (Bhamare et al., 2020).

The deployment and installation of Edge computing infrastructure are hampered by security concerns. If not integrated appropriately with current centralized authentication systems, it introduces malicious assaults, sporadic connection, and slowness. Few areas of network infrastructure have restricted access to authentication servers, and as a result, they will not be effectively integrated and implemented. Another significant problem in designing privacy-protection techniques is ensuring the user's location and service use are protected. At each tier of the edge cloud ecosystem, from edge devices to cloud servers, security should be addressed. (Uddin and Ahmad, 2020).

ENISA, the European Union Agency for Cybersecurity, has produced research on both general IoT security guidelines and sector-specific IoT security best practices. According to the research, while horizontal and vertical IoT security measures both contribute significantly to risk reduction, the design, development, implementation, and configuration of secure IoT systems should not be overlooked. ENISA aggressively advocates for security and privacy by design by default, stating that some of the IoT's security concerns may be solved by adopting a set of safe development rules. This involves vulnerability scanning, secure deployment, and guaranteeing the continuation of secure development. Software is a critical component of the whole IoT supply chain, as adversaries may attack the security of IoT devices and services by exploiting software flaws. When estimating risk, it is necessary to include the whole IoT ecosystem. IoT software development cannot be done in isolation from the underlying hardware. Elements such as the Root of Trust or the Chain of Trust demonstrate the relationship between software and hardware. (European Union Agency for Cybersecurity, 2019).

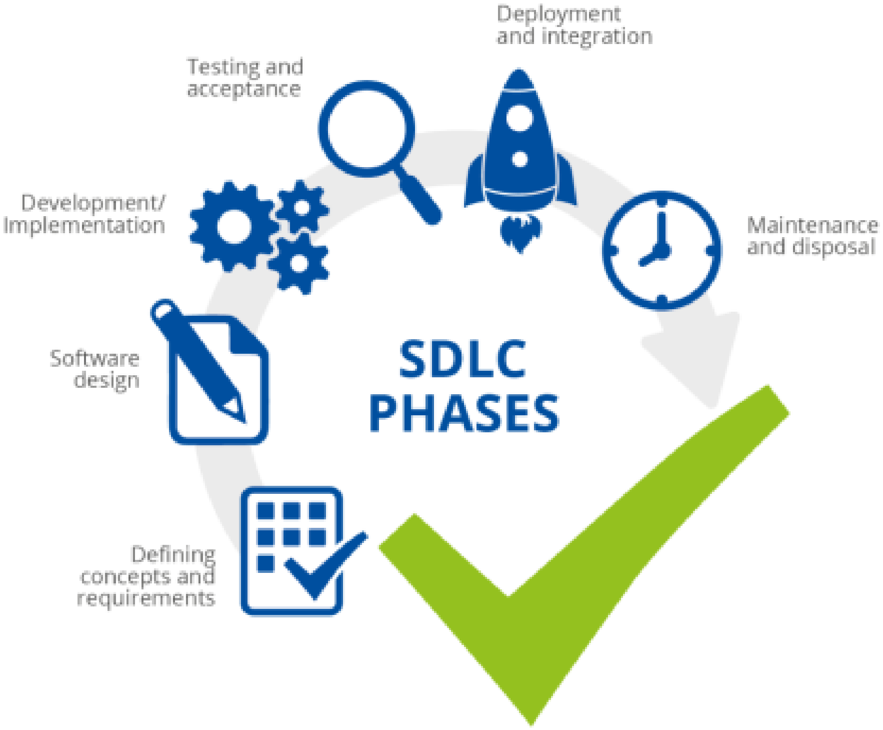

Deployment and integration
Testing and acceptance
Development/ Implementation
Maintenance and disposal
Software design
SDLC PHASES
Defining concepts and requirements

Figure 15: Software Development Life Cycle phases (European Union Agency for Cybersecurity, 2019).

The IoT Software Development Life Cycle (SDLC) consists of different phases that aim to deliver effective and efficient systems. There are several approaches to accomplish this target, as seen by the multiple SDLC models specified by ENISA. By addressing security carefully across the IoT SDLC and implementing suitable security measures on any impacted assets, the overall security of the IoT ecosystem is enhanced. (European Union Agency for Cybersecurity, 2019).

According to a security framework by Industrial Internet Consortium (2016), numerous stakeholders from various sectors are shaping the IIoT, all of which must take security into account. Usually, an IIoT system connects and integrates ICSs with enterprise systems, business processes, and analytics. Typically, these are systems that interact with the real world, where uncontrolled change might result in dangerous situations. Additionally, these IIoT systems may include data flows that involve several intermediate entities, necessitating security measures beyond basic connection encryption. IIoT systems, which have a lengthy life expectancy, contain installations that could be described as legacy.

The security framework defines and illustrates the processes for identifying, evaluating, and mitigating risks associated with security and privacy concerns. "*Security, safety, reliability,* *resilience, and privacy are the five attributes that have the most impact on an IIoT deployment's trust choices*," according to the framework.

Historically, the security of Operational Technology (OT) and Information Technology (IT) systems has been assessed separately. Trustworthy IIoT systems need an end-to-end evaluation of their security operations. "*To be successful, security, regulations, and standards must evolve*," the framework states. (Industrial Internet Consortium, 2016).

To continue further on the subject, Gartner (2021d) report state "*there is a rise in the frequency of attacks against operational technology (OT) - the hardware and software that monitors or controls equipment, assets, and processes.*" Additionally, they have progressed beyond causing immediate disruption to operations, such as shutting down a factory to endangering the integrity of industrial ecosystems in order to inflict bodily injury. (Gartner, Inc., 2021d).

According to the Gartner report, security incidents in OT and other CPS systems are primarily driven by three factors: actual damage, economic vandalism (decreased productivity), and reputational vandalism (making a manufacturer untrusted or unreliable).

The Gartner report forecasts the economic effect of lethal CPS attacks would top $50 billion by 2023. Even if one disregards the fundamental worth of human life, the costs to businesses in terms of compensation, litigation, insurance, regulatory fines, and reputational harm would be severe. The Gartner report recommends businesses adopt a framework of 10 security principles to help them boost their security posture across their facilities and minimize adverse effects of digital events on the physical world. (Gartner, Inc., 2021d).

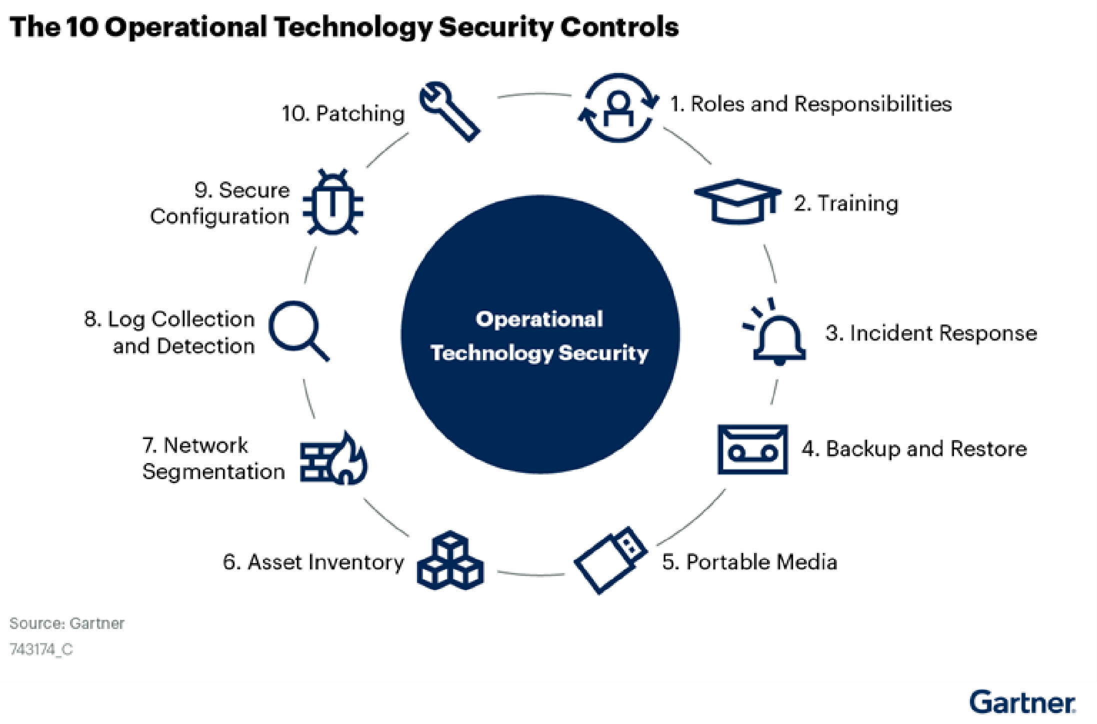


Figure 16: The 10 Operational Technology Security Controls (Reprinted, with permission, from Gartner, Inc,. 2021d © Gartner, Inc.).

1. Define roles and duties precisely: Each institution requires an operational technology security manager. Staff, senior management, and third parties shall be assigned and documented security roles.

2. Provide enough awareness and training: Every OT professional must possess the necessary skills to perform their responsibilities. Employees must be taught to identify security threats, attack vectors, and the proper course of action if a security breach occurs.

3. Develop and practice incident response: Preparation, detection and analysis, containment, eradication and recovery, and post-incident operations are all steps of addressing OT-specific security issues.

4. Disaster recovery: Ascertain the existence of sufficient backup, restore, and disaster recovery methods. Move the backup media off-site. Additionally, backup media must be safeguarded against illegal disclosure or use. The backup must be recoverable on a new system or virtual machine in difficult situations.

5. Manage portable media: Enforce scanning of all portable storage devices, like USB sticks, whether owned by internal employees or outside parties such as subcontractors or

representatives of equipment manufacturers. Only virus- and malware-free media are permitted to be connected to the OT.

6. Keep a precise inventory of OT hardware and software.

7. Segment networks correctly: Physical or logical separation of OT networks from other networks is required. A secure gateway solution, such as a demilitarized zone (DMZ), must be used between an OT and the rest of the network. Multi-factor authentication is required for interactive OT connections.

8. Records management and real-time detection: Automatic logging and assessment of potential and actual security incidents must be introduced, including security log retention durations and anti-tampering measures.

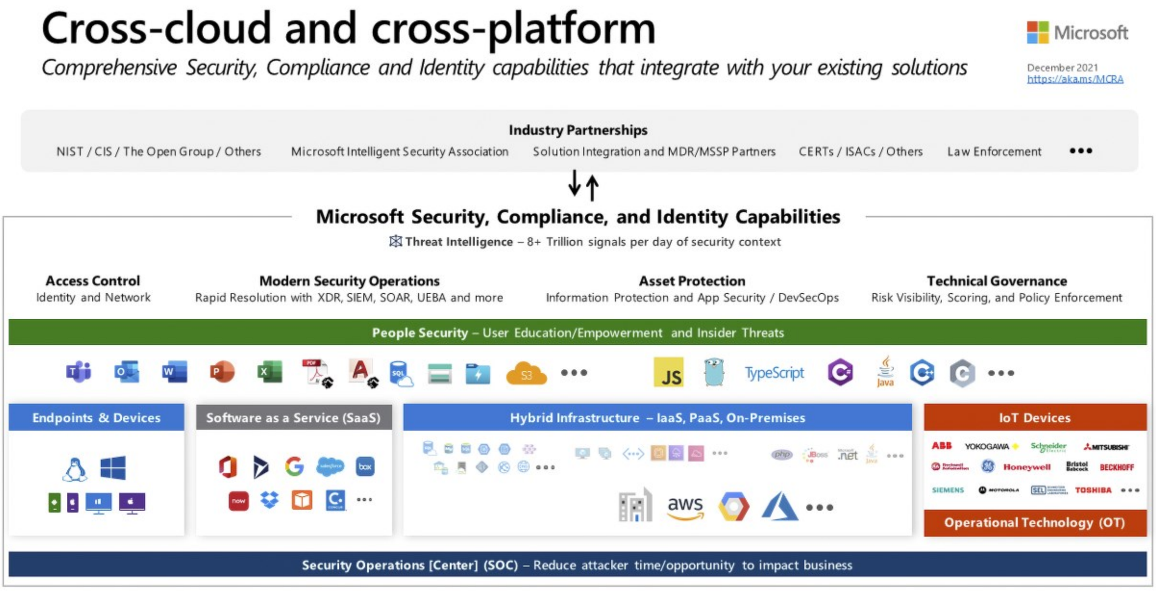
9. Defining a safe configuration technique: Endpoints, servers, network devices, and field devices must all have secure settings created, standardized, and applied. All supporting components must have anti-malware software installed and active.

10. Formal patching procedure: Establish a process for testing patches with equipment suppliers. Once validated, the patches may be deployed only to selected systems and at specific intervals. (Gartner, Inc., 2021d).

One recent example of a zero-day vulnerability that got the whole internet by surprise is the famous Java logging library log4j, used in almost all major Java-based enterprise apps. According to Lunasec, "*given how ubiquitous this library is, the impact of the exploit, and how easy it is to exploit, the impact of this vulnerability is quite severe.*" (Wortley and Thompson, 2021).

To summarize the situation, security is getting more and more complex because the battlefronts are evolving and changing all the time. Both Cloud and edge computing have their own characteristics, differences, and similarities in terms of security.

2.6.2 Key factors in security in Azure platform



Cross-cloud and cross-platform

Comprehensive Security, Compliance and Identity capabilities that integrate with your existing solutions

Microsoft

December 2021
https://aka.ms/MCRA

Industry Partnerships

NIST / CIS / The Open Group / Others | Microsoft Intelligent Security Association | Solution Integration and MDR/MSSP Partners | CERTs / ISACs / Others | Law Enforcement | •••

Microsoft Security, Compliance, and Identity Capabilities

Threat Intelligence – 8+ Trillion signals per day of security context

Access Control
Identity and Network

Modern Security Operations
Rapid Resolution with XDR, SIEM, SOAR, UEBA and more

Asset Protection
Information Protection and App Security / DevSecOps

Technical Governance
Risk Visibility, Scoring, and Policy Enforcement

People Security – User Education/Empowerment and Insider Threats

Endpoints & Devices

Software as a Service (SaaS)

Hybrid Infrastructure – IaaS, PaaS, On-Premises

IoT Devices

Operational Technology (OT)

Security Operations [Center] (SOC) – Reduce attacker time/opportunity to impact business

Figure 17: Microsoft cross-platform security diagram (Used with permission, Microsoft, 2021c).

Today, most organizations operate in a complex environment that includes numerous OS platforms, SaaS cloud services, and IaaS/PaaS cloud providers. Microsoft builds security for the multi-cloud and cross-platform enterprises, not just for Microsoft. (Microsoft, 2021c).

The fundamental concepts behind a Zero Trust security model

According to Microsoft (2021e), the three critical concepts behind a Zero Trust security strategy are as follows:

- *"Verify explicitly: Always authenticate and authorize based on all available data points, including user identity, location, device health, service or workload, data classification, and anomalies.*
- *Use least-privileged access: Limit user access with just-in-time and just-enough-access (JIT/JEA), risk-based adaptive polices, and data protection to help secure both data and productivity.*
- *Assume breach: Minimize blast radius for breaches and prevent lateral movement by segmenting access by network, user, devices, and app awareness. Verify all sessions are encrypted end to end. Use analytics to get visibility, drive threat detection, and improve defenses."*

Security layers of Zero Trust

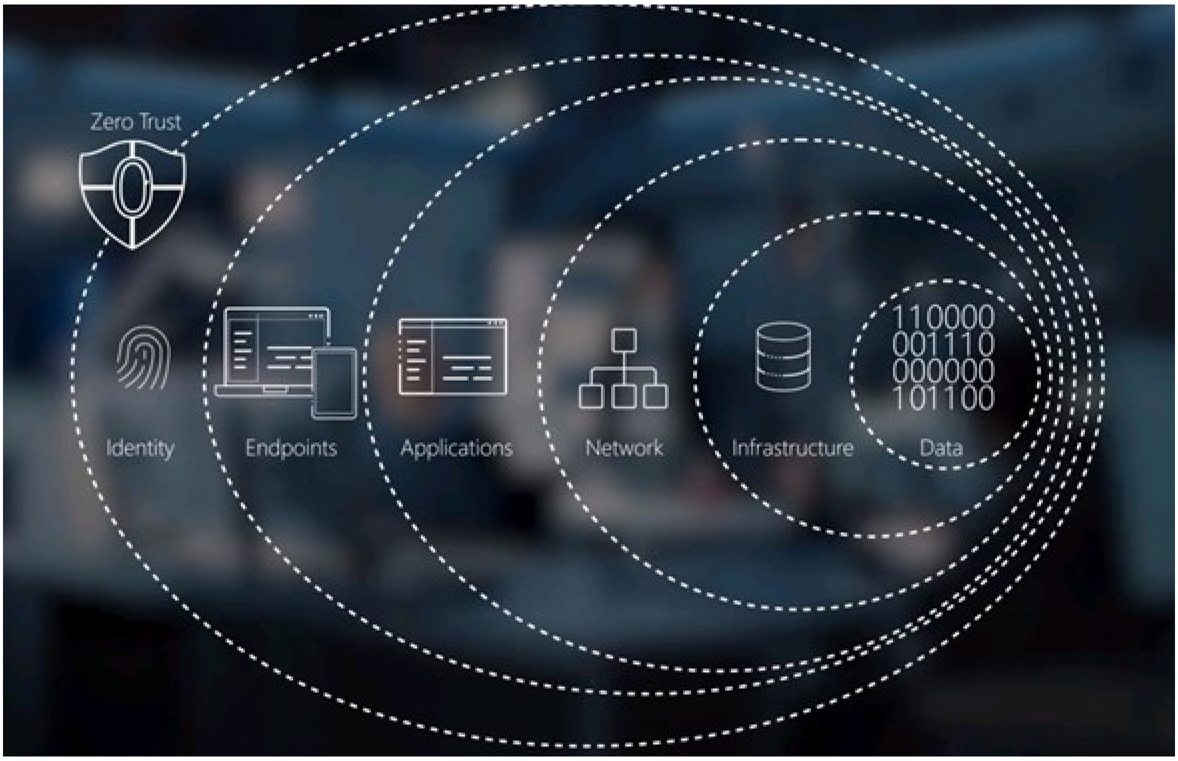


Figure 18: Zero Trust diagram (Used with permission, Microsoft, 2021f).

Microsoft's Zero Trust strategy is comprised on six distinct layers: Identity, Endpoints, Applications, Network, Infrastructure, and Data. These layers guarantee that the resources are accessible only to the people, devices, and processes that have been allowed access.

Identity

Azure Active Directory is a cloud identity service, which provides identity and conditional access restrictions to users, app and process service accounts, and devices. Also, Azure AD can offer a single identity control plane with standard authentication and authorization capabilities for any applications and services, not only Microsoft services. This avoids the usage of numerous credentials and weak passwords across services and helps provide passwordless multi-factor authentication for users using biometrics or a FIDO2 key. To make the authentication process less irritating for users, Conditional Access in Azure AD provides real-time intelligence upon sign-in. (Microsoft, 2021f).

Conditional access enables risk assessment. For example, an app may make sign-in location choices and enforce access restrictions in real-time to either block access and demand a password reset, permit access but require an extra authentication factor, or limit it to view-only rights. (Microsoft, 2021f).

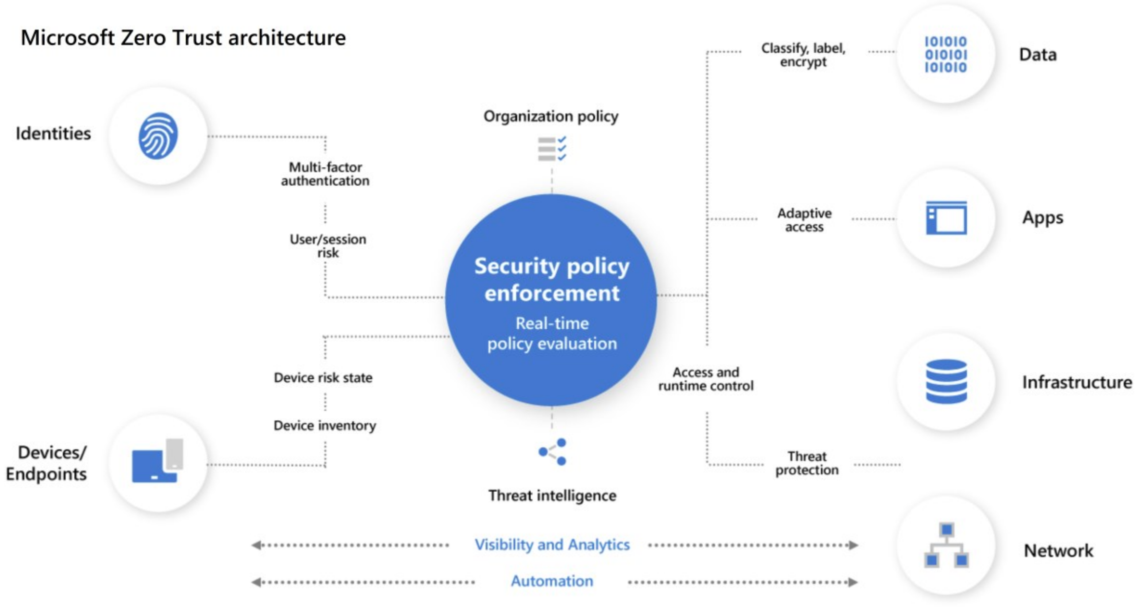


Figure 19: Microsoft Zero Trust architecture (Used with permission, Microsoft, 2021e).

Endpoints

The company may not own or manage the endpoints used by users to access resources. Endpoints that are not updated or secured risk data leakage from unknown applications or services. Microsoft Endpoint Manager ensures that devices and applications fulfill security and compliance policy requirements, whether owned by the enterprise or the user. This security applies whether the device connects from inside the network perimeter, over a VPN, a home network, or the public internet. (Microsoft, 2021f).

With its Extended Detection and Response (XDR) controls, Microsoft Defender can also detect and contain breaches identified on an endpoint, forcing the device back into a trustworthy condition before allowing it to reconnect to resources. (Microsoft, 2021f).

Applications

There are numerous approaches to safeguard apps using Zero Trust. As previously explained, using the Azure AD as a single entity provider for authentication and conditional access also apply to cloud-based services and local applications that connect to Azure AD. Microsoft Endpoint Manager allows managing policies for desktop and mobile programs, for example, preventing work-related data from being duplicated and utilized in personal applications. In order to mitigate emerging risks, organizations must know and decide which applications are utilized. (Microsoft, 2021f).

Network

Zero Trust security approach includes the network layer. Many controls exist for current architectures and hybrid services spanning on-premises and various cloud services, virtual networks, and VPNs. According to Microsoft (2021f), these controls include:

- Network segmentation to minimize attack radius and lateral movement
- Threat protection against DDoS and brute force attacks and the ability to identify and react promptly to incidents
- Encryption of all network communication, internal or external

Products like Azure Firewall and Azure DDoS Protection help safeguard Azure VNet resources. Microsoft XDR and SIEM products Microsoft Defender and Microsoft Sentinel help promptly detect and contain security incidents. (Microsoft, 2021f).

Infrastructure

It is crucial to managing configuration and software upgrades to ensure that all deployed infrastructure fulfills security and regulatory requirements. In Microsoft Azure, landing zones, blueprints, and policies guarantee that new infrastructure satisfies cloud resource compliance needs. Furthermore, Azure Security Center (which is being renamed to Microsoft Defender for Cloud) and Log Analytics also assist in managing configuration and software updates for on-premises, cross-cloud, and cross-platform infrastructure. (Microsoft, 2021f).

Security monitoring is essential for spotting threats and irregularities. Microsoft Defender with Microsoft Sentinel delivers multi-cloud threat prevention with automated detection and response. (Microsoft, 2021f).

Data

Ultimately, Zero Trust is about understanding and implementing effective data protection procedures. The Microsoft ecosystem provides controls to restrict data access to just those individuals and processes that require it. It is possible to limit or prevent undesired sharing of sensitive data and files by implementing policies. Microsoft Information Protection, for example, automates file and content tagging and categorization. Policies are then assigned to labels to perform protective measures like encryption, access control, and blocking third-party applications and services. (Microsoft, 2021f).

Azure Purview automatically identifies and maps data from Azure, on-premises, and SaaS sources and works with Microsoft Information Protection to identify sensitive information. (Microsoft, 2021f).

Zero Trust for Microsoft identity platform application developer

According to Microsoft (2022a), the Zero Trust model is a reassessment of security methods and procedures from the viewpoint of an implicit trust culture that must give way to an explicit verification culture. Incorporating the Zero Trust architecture into applications may boost security, decrease the blast radius of a security incident, and help in rapid recovery. A single app may be the weakest link in the chain, which hostile actors might exploit to get access to other business-critical data and processes.

Traditional network controls cannot be relied anymore on for security. Controls must be relocated to where data resides: on devices and inside applications. Microsoft's identity platform is intended to ease authentication and authorization for developers. It abstracts a significant amount of labor from the code a developer must write. There are specific steps that developers may take while using the platform that can influence the security of their applications. (Microsoft, 2022a).

According to Microsoft (2022a), these include for example:

- The rejection of outdated authentication protocols such as "Basic authentication" and adopting more current ones such as Open ID Connect and OAuth2.
- Azure Active Directory uses Conditional Access (CA) to aggregate signals, make access choices, and enforce corporate regulations. Conditional access may reroute the user to the identity provider for multi-factor authentication for enhanced security.
- Utilizing the Resource Owner's Password Credentials (username/password flow) is not recommended. This needs a great deal of trust in the application and introduces risks not seen in other flows.

- Utilizing application settings in the tenant to restrict access to the application.
- Utilizing the "User assignment needed" flag to restrict access to the application to a subset of users. To maximize flexibility and control, propose applications publish app roles and associate app roles with security groups.

The Microsoft identity platform application registration portal serves as the main entry point for apps that want to utilize the platform for authentication and related purposes. When registering and configuring applications, various decisions will impact the application's ability to adhere to Zero Trust principles. (Microsoft, 2022a).

Core Zero Trust capabilities for IoT

According to Microsoft (2021e), a breach in an organization's IoT security posture might have significant consequences such as:

- Impact on operations and income: IoT devices may experience operational degradation, be utilized for lateral movement, or be brought down as a consequence of a security event, resulting in revenue loss.
- Incidents may harm the consumer experience and harm a brand's reputation.
- Cyber-physical system (CPS) compromise may lead to real-world implications such as safety and environmental disasters.
- Non-compliance may have a detrimental effect on an organization's ability to comply with applicable government and industry regulations.

IoT solutions have several technical issues to consider, these include for example :

IoT devices are "userless" and execute scripts: Cameras, robots, and controllers are examples of "userless" IoT equipment. The "user" under the Zero Trust IoT concept is the device itself, which operates autonomously. Numerous applications operating on these edge devices are automated, containerized, and continuously running. (Microsoft, 2021e).

Platforms for IoT devices are diverse and interconnected: Many IoT installations use older devices and technology meant for a disconnected environment. Operating systems range from bare metal and real-time operating systems to sophisticated operating systems, with many lacking update capabilities and including insecure open source components. This means that devices might be vulnerable for considerably longer than the average worker's PC or smartphone. Many IoT devices are tiny and cannot run a complete OS stack, security agents, or encryption. Battery-powered devices have limited processing power and size. Networking topologies might impair the ability to manage and monitor devices and workloads. (Microsoft, 2021e).

IoT devices might be considered high-value targets. These devices, which are often utilized in mission-critical services and facilities, might be tempting targets for cybercriminals seeking command and control. The sheer quantity of IoT devices makes them attractive targets for botnets. For instance, the Mirai Botnet exploited IoT devices to disrupt internet service. (Microsoft, 2021e).

Physical or local threats might affect IoT devices. They are used both within and outside of protected organizational settings. For instance, a security camera may be exposed to direct physical attacks by adversaries. (Microsoft, 2021e).

Device authentication requires a strong identification

According to Microsoft (2021e), the first element of Zero Trust for IoT is strong device identification, which is achieved by closely integrating "*IoT devices and services including:*

- *A hardware root of trust.*
- *Password-less authentication.*
- *Renewable credentials.*
- *Organizational IoT device registry.*"

Devices' initial security property is a hardware root of trust. A strong IoT device identity begins with the hardware and manufacturing process. This identification is unique to the physical device and cannot be altered over the lifetime of the device. (Microsoft, 2021e).

Passwordless authentication uses standardized X.509 certificates to verify a device's identity. Its power stems from the fact that it uses a private key to sign and encrypt and a public key to verify and decrypt. This is more secure than passwords and symmetric tokens, which depend on the device and service sharing a common secret. (Microsoft, 2021e).

After registering with an IoT solution, a device must be frequently provided renewable credentials to ensure continued security. Renewable credentials provide a device's operating identification. Operational certificates must be issued by a trustworthy PKI and have a renewal period that matches the business's security posture. Certificate renewal must be automated to avoid manual rotation causing access issues. (Microsoft, 2021e).

Securing an organization's IoT devices requires a registry to manage their lifespan and audit device access. The information included in the IoT device registry is used to onboard devices into an IoT system by confirming their identity and credentials. Once onboarded, the device registry holds the device's operating identification and credentials for daily usage. The IoT device registry data may be queried and organized in order to facilitate devices for scalable operation, management, workload deployment, and access control according to Microsoft (2021e).

Access using the least privileges possible to reduce blast radius

When combined with the strong identity supplied by linked devices and services, Zero Trust necessitates least-privileged access control to minimize the impact of hacked identities or disapproved workloads. This entails controlling access to devices and their workloads through device and workload access control and Conditional access, according to Microsoft (2021e).

Managing device access is sufficient method to handle access to the device's scoped operations in the case of smaller, single-purpose IoT devices. When dealing with multiple parties, these workloads should have distinct application identities that, when paired with the device identity, allow access control and auditing. (Microsoft, 2021e).

Integrated identity and access management system should manage device and workload access via a single pane of glass. A device, an application, or a combination may be used to control and audit access. (Microsoft, 2021e).

Conditional access to devices and their workloads: The caller's operational circumstances must be examined prior to authorizing devices and workloads. These include criteria such as geographic location, uniqueness, and time signals. (Microsoft, 2021e).

Restrict access or flag devices for remediation using device health

In accordance with the Zero Trust concept of "verify explicitly," the health of a device should be a crucial aspect in establishing the risk profile of the device. This might be used as a way to verify that access to IoT services and apps is granted only to healthy devices. (Microsoft, 2021e).

Devices that are unhealthy may be isolated, inspected, and remedied in order to reclaim their health. Numerous features may be combined to provide a unified picture of device health. Those solutions include security configuration assessment, vulnerability assessment, insecure credential assessment, active threats and threat alerts, and anomalous behavioral alerts. (Microsoft, 2021e).

Updates to keep devices healthy

According to Microsoft (2021e), the flip side of allowing health-based device access control is proactive maintenance of production equipment in a functional, healthy condition. Among the capabilities that contribute to the health of devices are:

Device administrators must adopt a centralized configuration and compliance management strategy to establish a strong security posture for their devices and implement policies that enable them to comply with industry-specific compliance requirements. In addition to controlling settings, the centralized configuration must secure certificate distribution and update for IoT devices.

Deployable Updates: Keeping devices updated is vital to protecting them. Customers must update Cloud-deployed workloads and base OS. The update system should provide remote deployment and validation and, preferably, should be coupled with pervasive security monitoring to allow security fixes. Achieving device health as an access point and the desired condition for Zero Trust IoT requires integration across organizational units and job functions. (Microsoft, 2021e).

Monitoring and responding to security risks

According to Microsoft (2021e), security monitoring is required to detect unauthorized or compromised devices quickly. Monitoring protects managed "greenfield" devices and compensates for unmanaged "brownfield" devices that lack agents and are not easily patchable or configurable remotely.

CISA recommendations, as cited by Microsoft (2021e), include:

- *"Generating an "as-is" asset inventory and network map of all IoT/OT devices.*
- *Identifying all communication protocols used across IoT/OT networks.*
- *Cataloging all external connections to and from IoT/OT networks.*
- *Identifying vulnerabilities in IoT/OT devices and using a risk-based approach to mitigate them.*
- *Implementing a vigilant monitoring program with anomaly detection to detect malicious cyber tactics like "living off the land" techniques within IoT/OT systems, such as monitoring for unauthorized changes to controllers."*

The majority of IoT attacks follow a "kill chain" pattern: attackers get initial access, ensure persistence, escalate privileges, and then traverse the network laterally. They often use privileged credentials in order to overcome network segmentation obstacles such as next-generation firewalls. Detecting and reacting quickly to these multistage threats needs automation, machine learning, and threat intelligence across IT, IoT, and OT networks. Security operations center (SOC) analysts may search for and uncover previously unknown threats by using centralized SIEM and XDR systems. (Microsoft, 2021e).

Security Orchestration and Automated Response (SOAR) systems are critical for quickly reacting to crises and reducing attacks before they do significant damage. A good practice is to define playbooks that are performed automatically when particular incidents are recognized. Compromised devices may be automatically banned or isolated to prevent infection. (Microsoft, 2021e).

Azure IoT services security

The Azure IoT Hub Device Provisioning Service (DPS) allows organizations to register devices for bulk onboarding, acting as a central device registry. It supports device certificates for onboarding, identification, and credential renewal, as well as device registration for their operation in IoT Hub. Azure IoT Hub enables the management of an enterprise-wide IoT device registry and accepts device-specific certificates for the purpose of enabling strong identity. To prevent illegal connections, devices may be deactivated centrally from Azure IoT Hub. Azure IoT Hub enables module identities to be provisioned for IoT Edge applications. (Microsoft, 2021e).

Azure Device Update (ADU) provides over-the-air (OTA) upgrades to IoT devices. It is a cloud-based system that enables the connection of nearly any device. ADU is extendable through open source and supports Linux and Azure RTOS (Real-Time Operating System) among other IoT operating systems. (Microsoft, 2021e).

Microsoft Sentinel (before rebranding known as Azure Sentinel) is a cloud-native SIEM/SOAR platform that provides a 360-degree view of corporate security. Sentinel can identify unusual behaviors indicative of IoT/OT device compromise by gathering data from all users, devices, apps, and infrastructure. Microsoft Sentinel accelerates threat detection and response by using machine learning. Furthermore, it offers all of the advantages of a cloud-based service, such as ease of use and flexibility. (Microsoft, 2021e).

Microsoft Defender for IoT (before rebranding known as Azure Defender for IoT) is an agentless, network-layer security platform for IoT and OT systems that enables continuous asset discovery, vulnerability management, and threat detection. The system employs unique IoT/OT-aware behavioral analytics and threat intelligence to identify unusual or illegal activity rapidly. (Microsoft, 2021e).

It enables quick deployment due to the fact that it does not need any modifications to current devices. It is compatible with a broad range of industrial equipment and protocols from the leading companies in the field of OT automation. Also, it integrates with Microsoft Sentinel and other SOC solutions. Interoperability with Microsoft Sentinel provides speedy detection and automatic response to multistage assaults that often span IT/OT boundaries. Microsoft Defender for IoT is available in a variety of deployment models, including fully on-premises, cloud-connected, and hybrid. (Microsoft, 2021e).

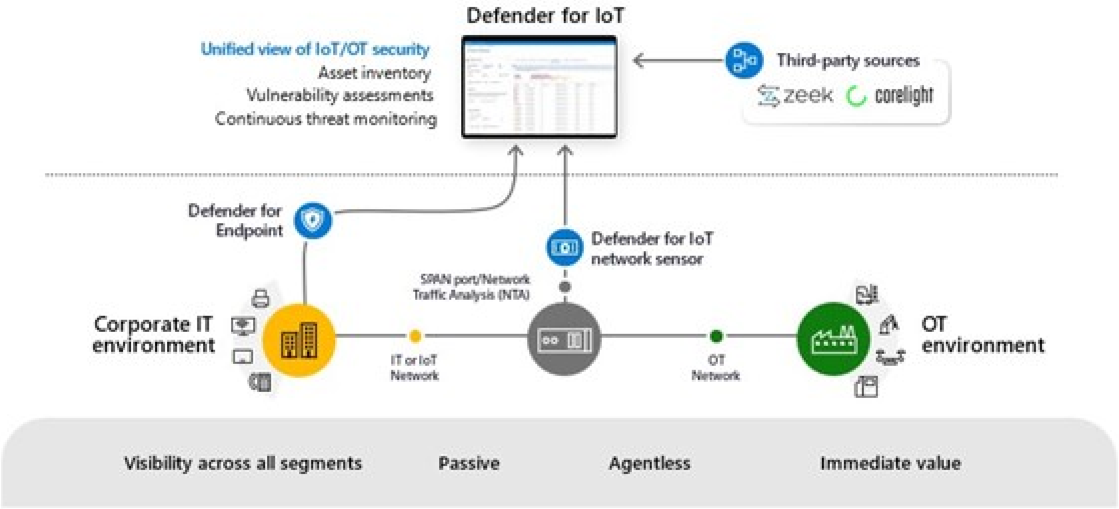

Defender for IoT
Unified view of IoT/OT security
Asset inventory
Vulnerability assessments
Continuous threat monitoring
Third-party sources
zeek
corelight
Defender for Endpoint
Defender for IoT network sensor
SPAN port/Network Traffic Analysis (NTA)
Corporate IT environment
IT or IoT Network
OT Network
OT environment
Visibility across all segments
Passive
Agentless
Immediate value

Figure 20: Microsoft Defender for IoT (Used with permission, Microsoft, 2021e).

Additionally, Microsoft Defender for IoT includes lightweight agents that enable developers of IoT/OT devices to include robust security at the device layer, as well as endpoint detection and response (EDR) capabilities. These micro-agents provide source code for IoT operating systems like Linux and Azure RTOS. The agents connect with the Microsoft Defender for IoT console on the Azure portal. (Microsoft, 2021e).

Azure IoT Edge and Device Security

According to Jensen (2019), physical access is one of the primary attack surfaces unique to IoT Edge solutions in comparison to a standard cloud solution. The IoT Edge devices are not located in a data center. Furthermore, since these devices are so near to the source of most data for IoT Edge solutions, a breach affects anything that uses their data. As a result, they must be protected in such a manner that the data they generate can be trusted.

The commonality of the hardware in IoT Edge solutions creates a second security vulnerability. Due to economic concerns, most edge servers use commodity hardware. This implies that the methods used to attack the edge device are widespread and readily accessible simply because consumer-grade components adhere to hardware standards. (Jensen, 2019).

As a result, attackers often possess a thorough grasp of how to exploit edge server hardware. According to Jensen (2019), "*they may already be equipped with all the required tools to connect to and compromise an edge device.*"

The asset's worth is also a consideration in edge solutions. While IoT solutions and IoT edge devices share some hazards, they lack the computer intellectual property. Additionally, when more sophisticated analytics and specialized corporate modules containing years of proprietary knowledge are deployed on edge servers, they will become increasingly desirable to both the asset owner and prospective attackers. This creates a unique circumstance in which the IoT Edge is the perfectly accessible, high-value target, according to Jensen (2019).

Microsoft provides a security framework for the Azure IoT Edge that lays the groundwork for safe device deployment. The architecture is secured using secure silicon, which acts as a tamper-resistant root of trust. Secure execution environments dependent on adequate authentication, authorization, and attestation are crucial in the intermediate layers of the foundation. Application runtime integrity monitoring completes the framework's base. (Stackowiak, 2019).

According to Stackowiak (2019), at its most fundamental level, a best practice is to describe devices that fulfill the fewest possible physical requirements. If components like USB ports are deemed optional, they should be avoided since they potentially expose the device to intrusion. Secure enclosures and other methods may safeguard devices against physical manipulation. In device processors, a Software Guard Extension (SGX) may ensure that regular programs run in impenetrable enclaves. Microsoft's Azure IoT Edge Security Manager ensures the device's security even if the operating system is hacked. According to Stackowiak (2019), it is accountable for:

- *“Secure and measured device booting*
- *Device identity provisioning and trust transfer*
- *Hosting and protection of the Device Provisioning Service*
- *Securely provisioning IoT Edge modules with unique identities*
- *Serving as a gatekeeper to a device hardware root of trust*
- *Monitoring the integrity of IoT Edge operations at runtime“*

Securing the device's operating system via systems resource access control and privileged operations provides additional security. The Azure IoT Edge software's runtime integrity monitoring helps secure the device's overall computing environment. (Stackowiak, 2019).

Authentication is used in the foundation to ensure that access is restricted to just trusted parties, modules, and devices. Certificate-based authentication is the principal method of authentication and is derived from Internet Engineering Task Force (IETF) standards controlling public key infrastructure (PKI). (Stackowiak, 2019).

Authorization relates to the scope of permissions provided to devices, modules, and actors. It is often set with the fewest privileges possible, granting just the data and resources necessary to provide the desired business solution. In some instances, an authorization may be controlled through role-based access control (RBAC) or certificate signing privileges. (Stackowiak, 2019).

Attestation ensures program integrity throughout device startup, runtime, and software upgrades. According to Stackowiak, "*secure boot-up ensures the integrity of all software on the device, including the operating system, runtimes, and configuration data*." With device and security framework countermeasures, runtime attestation detects malware injections, unauthorized physical access, and configuration modifications. It ensures safe software patching and updates by measuring and signing packages. (Stackowiak, 2019).

Azure RTOS and integrated IoT hardware platforms

Azure RTOS is an embedded and IoT RTOS that is offered as C-language libraries. Because Azure RTOS apps may run on a variety of platforms, customers are responsible for their security. However, Azure RTOS is intended to work with Azure Defender for IoT, DPS, and Azure IoT Hub. It shares several security features with bigger, more costly IoT devices. (Microsoft, 2021e).

According to Microsoft (2021e), "*Azure RTOS provides support for Zero Trust design on microcontroller platforms that support hardware security features*." Some examples of security designs that integrate firmware and hardware include the ARM TrustZone, secure element devices, and industry standards like the ARM Platform Security Architecture. Azure RTOS users may develop Zero Trust designs for even the tiniest embedded IoT devices by combining these secure elements. (Microsoft, 2021e).

Azure Sphere

Three technological components work together to construct the Azure Sphere platform: security service, operating system, and a novel protected microchip.

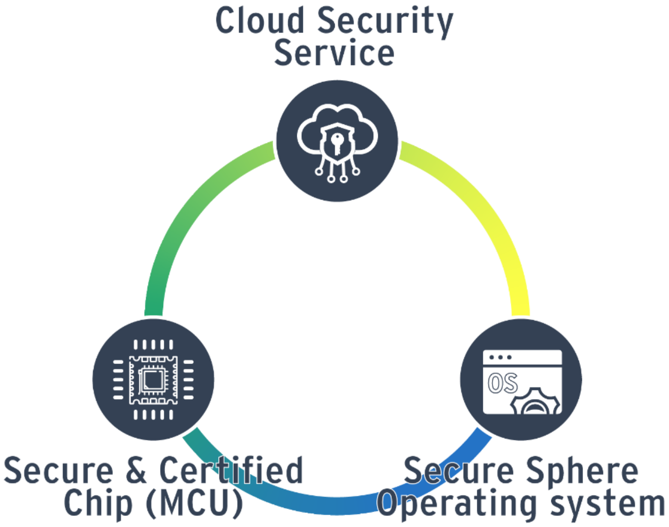


Figure 21: Azure Sphere platform (Adapted, with permission, from Bansal, 2020 © Springer).

Currently, the importance of incorporating each low-cost network-connected IoT device with high-value security is being underrated. Due to the IoT system's enormous numbers of end devices, device provisioning, configuration, and administration are challenging to integrate safely and effectively. According to Shi, Jin, and Li (2019), this connection may continue to be cost-effective by using Azure cloud services and Azure Sphere, which is all seven of the highly secure device criteria satisfying, fully managed integrated software, hardware, and cloud platform solution. It is based on the Zero Trust concept, which requires it to assume breach, even those involving its own software applications and operating system. Throughout the OS's architecture, protections are built with defense in depth. (Microsoft, 2021e).

For the Azure Secured Sphere OS, Microsoft's Linux-based microcontroller operating system serves as the foundation for new secure and agile IoT experiences. Although it is based on mainline Linux 4.9, a modified version of the Linux kernel created by Microsoft is utilized. Each device is equipped with a certificate and is known to Microsoft prior to the sale of Azure Sphere. As a result, Microsoft maintains and upgrades the operating system through over-the-air (OTA). Any cloud-based device management layer, including Azure as a preferred cloud, may be connected to Sphere. (Bansal, 2020).

2.7 Evaluation of Platforms in a research environment – Case VTT

Case Company VTT was established in 1942 in Finland and has been at the heart of resolving global problems via the use of science and technology. VTT's 80-year history demonstrates how research, technology, and collaboration can transform even the most significant societal concerns into possibilities for prosperity. VTT is a limited liability corporation owned entirely by the Finnish state and managed by the Ministry of Economic Affairs and Employment. VTT organizes its research, development, and innovation operations around three business areas: "*carbon neutral solutions, sustainable products and materials, and digital technologies*." (VTT Technical Research Center of Finland Ltd., 2021).

Even after the author’s rigorous attemps to promote the survey, only 28 responses were received, which weakens the survey results' reliability as a quantitative study. One probable explanation for the survey's low response rate was its timeframe. The pace of the end-of-the-year time is often frenetic, with individuals concentrating on their job and ignoring external irritations. The author personally received over a dozen email notifications requesting to complete different surveys, so one can only imagine how many distracting emails the researchers would have received throughout the same period.

However, the research data collected provides indications for reflection on the subject.

Survey results

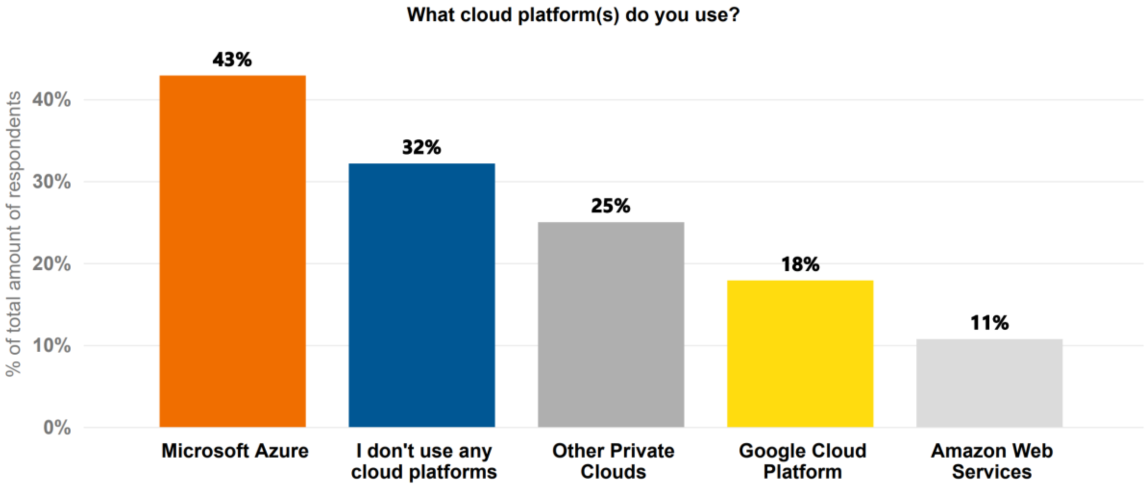


Figure 22: Survey Question "What cloud platform(s) do you use" (n=28)

As seen in Figure 22, the most used cloud platform amongst the respondents was Microsoft Azure, with 43% of total answers. Surprisingly, with 32%, the option "I do not use any cloud platforms" was the runner-up. Private clouds usage is widespread at VTT, so 25% was not an unexpected result. Google Cloud Platform and Amazon Web Services got 18% and 11 % share of the answers, respectively.

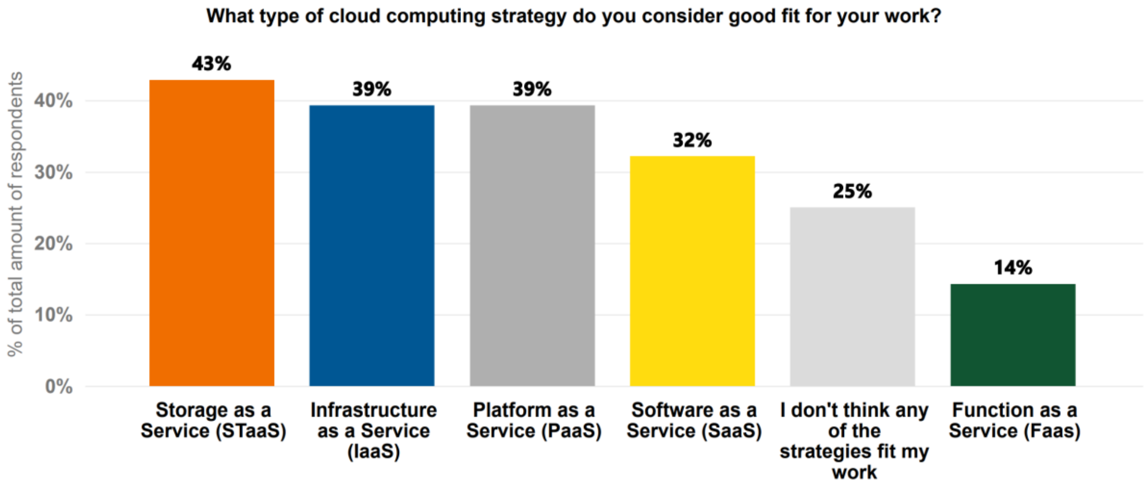


Figure 23: Survey Question "What type of cloud computing strategy do you consider good fit for your work" (n=28)

Storage as a Service was considered the best strategy for respondents' work, which is a suspected result since research is based on data. All of the most frequently utilized cloud service models received almost the same proportion of responses. Function as a Service was considered a good fit by only 14% of respondents.

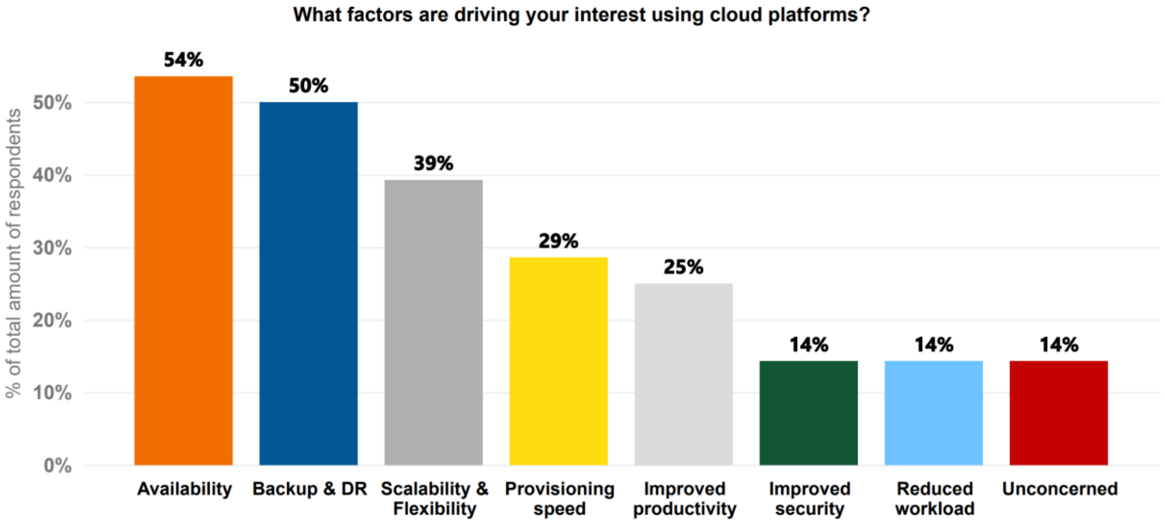


Figure 24: Survey Question "What factors are driving your interest using cloud platforms" (n=28)

The majority of the respondents answered that availability is the most significant factor, data backup and disaster recovery being the second. Scalability, development speed, and improved productivity settled between 25-39% of the answers. Improved security options and reduced workload were surprisingly the least exciting drivers.

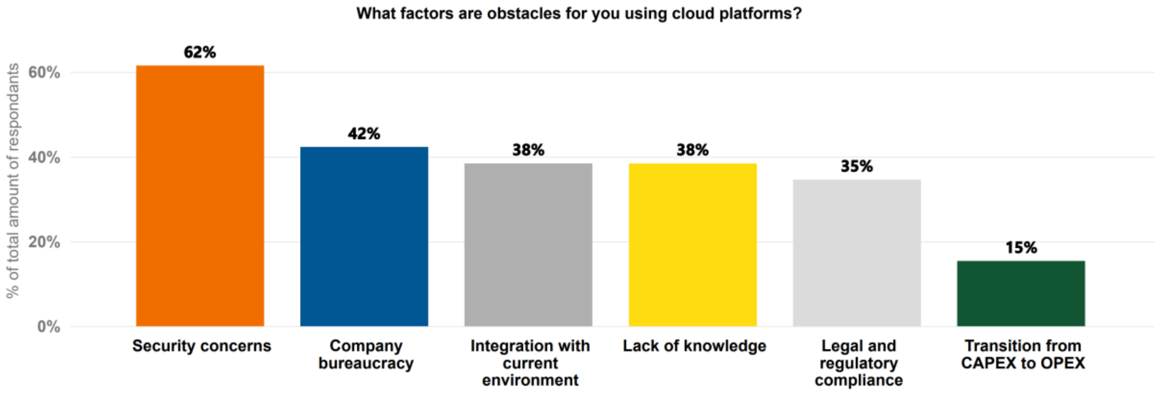


Figure 25: Survey Question "What factors are obstacles for you using cloud platforms" (n=28)

As seen in Figure 25, the main obstacle in using cloud platforms amongst the respondents was Security concerns, with almost two-thirds of the total answers. All the other options, excluding CAPEX to OPEX transition, were roughly on the same level. The answers give good insights into what issues matter to the researchers.

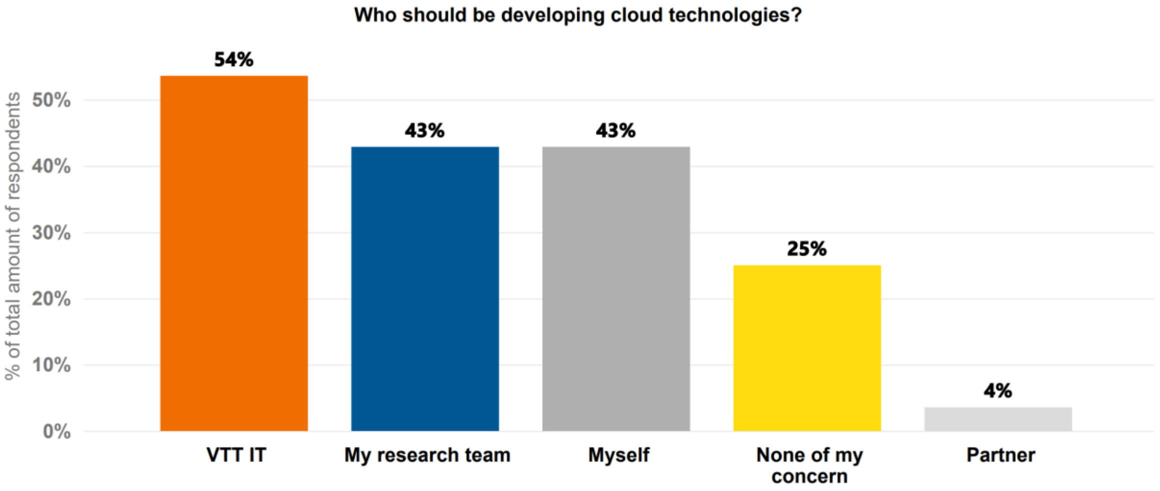


Figure 26: Survey Question "Who should be developing cloud technologies" (n=28)

VTT IT was considered by most for developing cloud technologies, with options "My research team" and "Myself" coming next. This is a suspected result for at least two reasons; first, the curiosity is built-in blood of VTT DNA, and second, researchers expect and trust in-house IT to provide service. Partner as a developing option was considered by merely 4% of respondents.

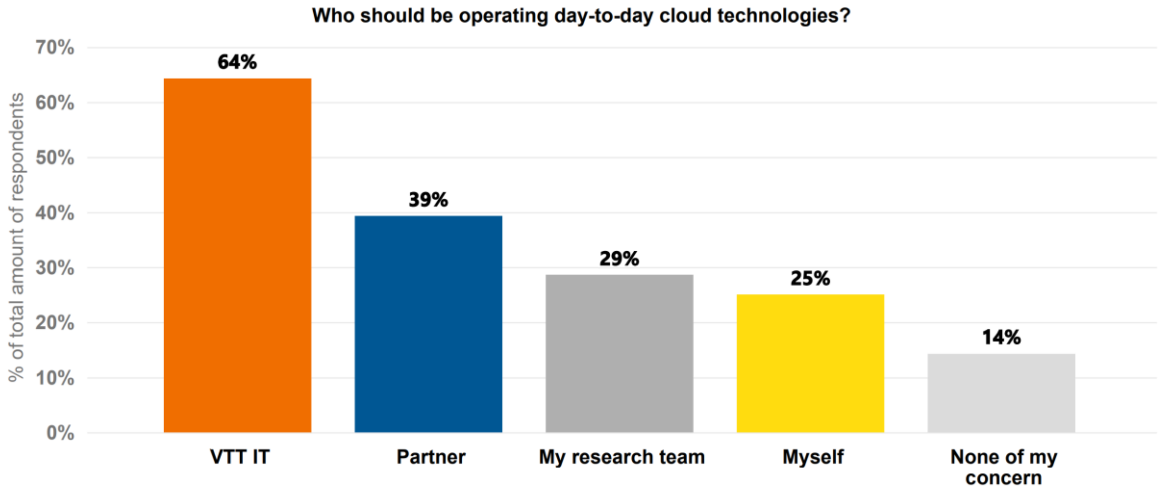


Figure 27: Survey Question "Who should be operating day-to-day cloud technologies" (n=28)

The shift from researchers to service providers such as VTT IT and partners can be seen in the answers when moving from development to day-to-day operations. The majority of the respondents answered that VTT IT is the one who should be in charge of operations. One-seventh of the answers were in the "None of my concern" category, which could mean the current state of the technologies operating is considered as normal as water, heat, and electricity, until they fail.

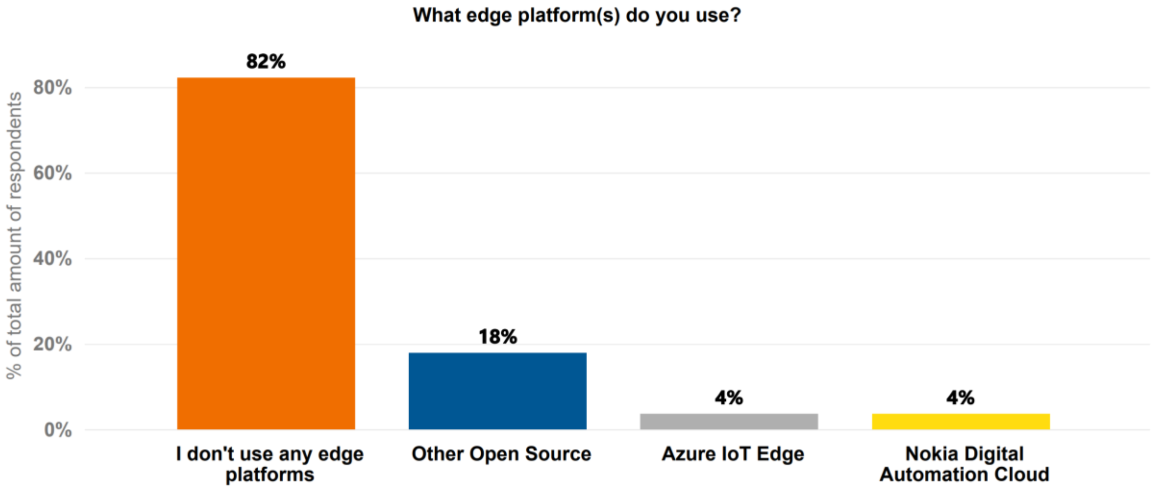


Figure 28: Survey Question "What edge platform(s) do you use" (n=28)

Considering the staggering percentage of option "I don't use any edge platforms", it can be assumed the survey did not reach the hoped target audience. Only a handful of responses stating Open Source platforms added with Azure IoT Edge and Nokia Digital Automation Cloud were used.

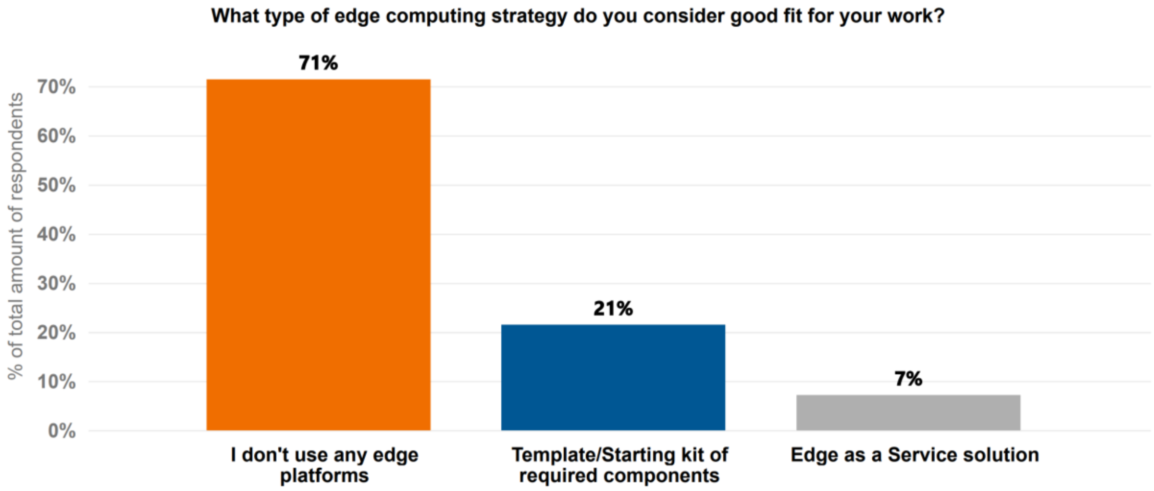


Figure 29: Survey Question "What type of edge computing strategy do you consider good fit for your work" (n=28)

Moving on from the previous question, when considering edge computing strategy fit for work, Figure 29 shows there is a need for templates and starting kits of required components. This is an important signal for VTT IT to start co-operating with researchers in that field to make it easier for others to start learning with ready-made starting kits and eventually make a usable Edge as a Service solution.

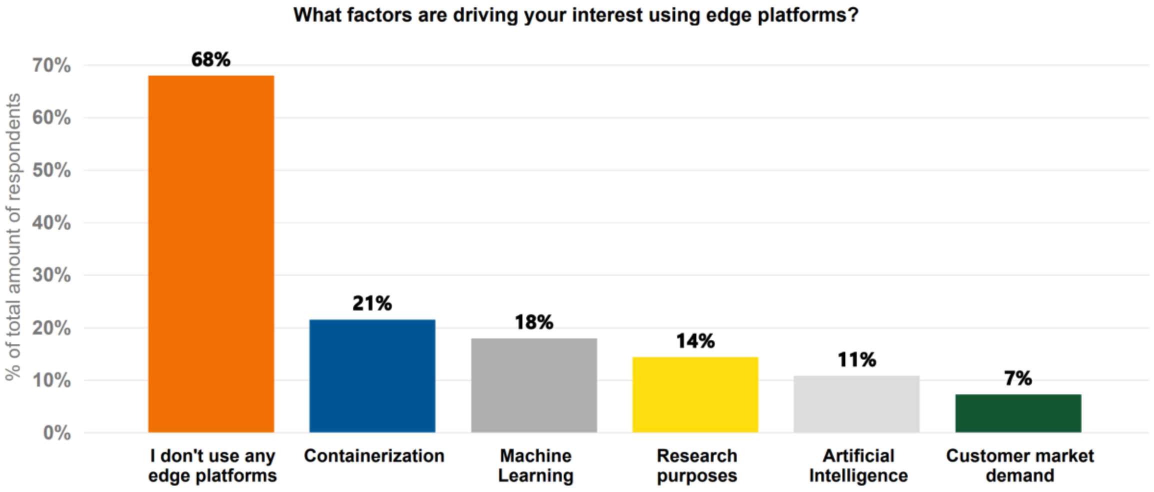


Figure 30: Survey Question "What factors are driving your interest using edge platforms" (n=28)

As seen in Figure 30, the main interests using edge platforms like containerization, machine learning, research purposes, and AI. The percentage of answers in option "Customer market demand" leaves much to speculate is it a real-life situation or a result of missing the target audience with the survey. Based on the publicly available references, the customer market demand is accelerating as companies learn how to benefit from the edge platform advantages and make their business more profitable.

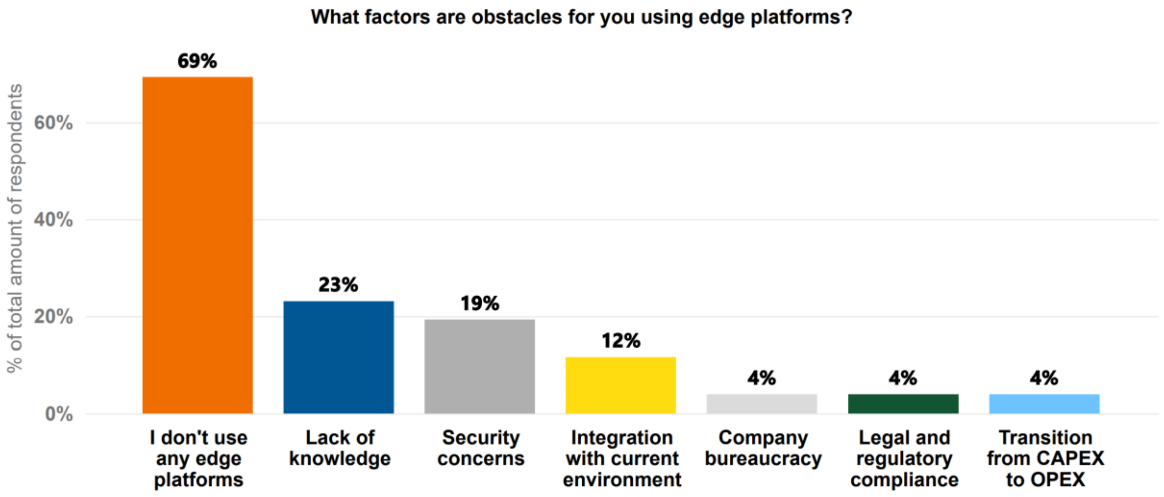


Figure 31: Survey Question "What factors are obstacles for you using edge platforms" (n=28)

Excluding the majority of the respondents who answered they are not using any edge platforms, lack of knowledge and security concerns were the two top selections. An interesting detail between cloud and edge platform questions is the options "legal and regulatory compliance" and "company bureaucracy" difference. In edge, only 4% answered they were obstacles. Compared to the cloud, a whopping 35% on option "legal and regulatory compliance" and 42% on option "company bureaucracy" stated they were obstacles.

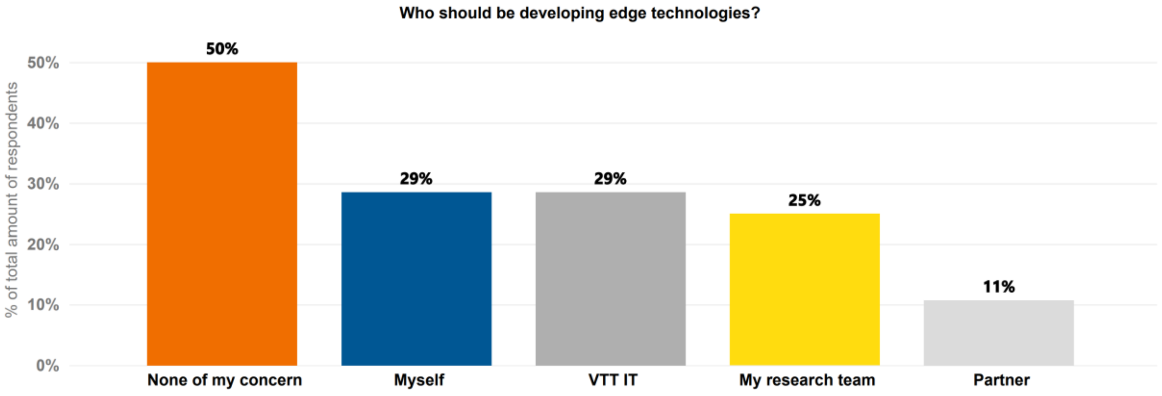


Figure 32: Survey Question "Who should be developing edge technologies" (n=28)

The majority of the respondents answered they are not using any edge platforms. After that, the distribution of answers is roughly the same between single individuals, VTT IT, and research teams. The "Partner" answer percentage was higher than in the cloud counterpart.

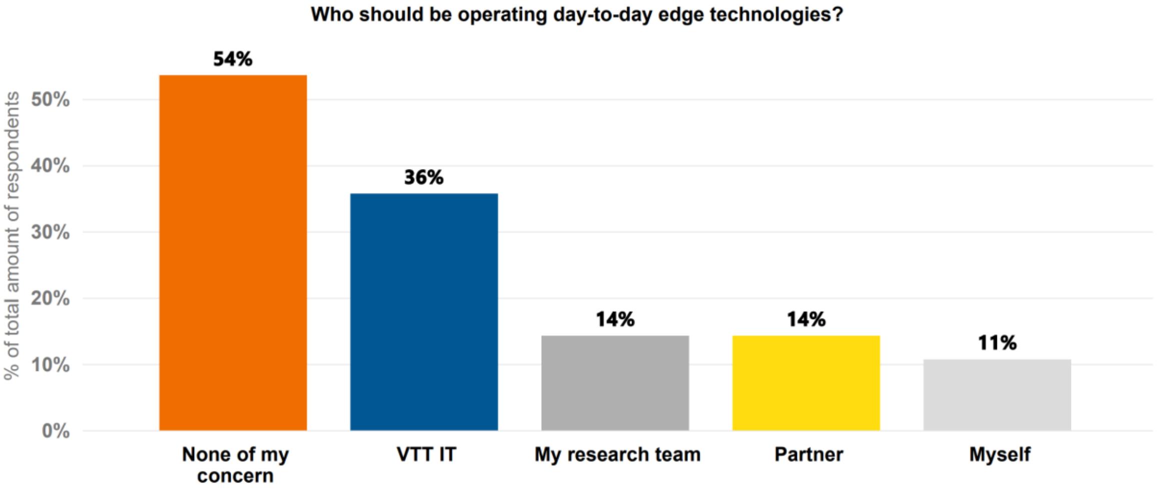


Figure 33: Survey Question "Who should be operating day-to-day edge technologies" (n=28)

Like in the equivalent cloud question, the shift from researchers to VTT IT can be seen in the answers when moving from development to day-to-day operations.

In addition to the previously visualized answers to the multiple-choice questions, the survey contained three free-form questions. The responses held some essential feedback to VTT IT, summarized in the following section.

Question 1: Is there anything else you would like to tell us about the adoption or use of cloud computing in your work:

1. Education and learning curve: Unleashing the full potential of the cloud platforms takes quite a bit of learning and skill to get it up and running.

2. Linux support: Research needs are mostly happening on Linux.

3. Lifecycle management: From prototyping to production and what to do after the project ends?

4. Data management: More general understanding is needed of how to handle data (e.g., regulations, GDPR). More support is needed from legal and IT/Cyber Security.

5. Cost optimization: There is no point in paying more than needed to get the job done.

Question 2: Is there anything else you would like to tell us about the adoption or use of edge computing in your work:

1. The importance of open source: Ensuring interoperability and ease of use was considered a critical issue.

2. Edge is a research topic in itself: "Hands-on" developing solutions compared to commercial alternatives is essential since the edge is an emerging paradigm.

Question 3: Please share your ideas on how VTT IT could support you in your work better:

1. Balance between governance and researcher's freedom: Sometimes, strict or unclear company governance policies slow down or limit the research process.

2. Easier and faster access to cloud services: Setting up and developing different pre-built startup packages or templates for research scenarios would speed the process. Also, common tooling guidelines should be in place.

3. Operations support: Outsourcing (research) IT support to external companies is generally seen as a wrong approach; co-operation with in-house staff is considered a better option.

4. Cost management: Project service portfolio costs after the project ends are difficult to handle. Also, the idea of the "pay-what-you-use" approach would give transparency to the costs.

5. Communication and education: Instructions, training courses, or example cases, and communicating these were considered crucial issues for success.

6. Data sovereignty issues: Having a clear and communicated solution for things like sensitive data processing, export control, or other legal issues regarding data was considered critical.

3 RESULTS

The survey's objective was to elicit responses on researchers' perceptions of the present cloud and edge situation at case study company VTT.

Twelve mandatory multiple-choice questions and three optional free-form questions were included in the anonymous survey.

The survey's intended audience was any cloud or edge platforms or technologies user. However, given the low response rate, this objective was not realized.

In retrospect, one reason the target was not accomplished might have been the survey's timeliness; most likely, the questions were either too specific or too broad or were deemed unworthy of time. For instance, the author may easily assert that the word "Edge" is unfamiliar to the majority of people or that it implies something different to various individuals.

Even with these low numbers, the findings provide insight and input on how VTT IT can improve the service's quality. Having open communication is a vital first step.

Microsoft Azure was the most often mentioned cloud platform by respondents. Other major cloud service providers were also used, which might indicate the possibility of multi-cloud servitization for a more robust cloud platform portfolio.

Storage as a Service was deemed the most appropriate service by respondents, which makes sense given that many respondents are unsure how to preserve results or the cloud environment for future project usage. Such an example is how to archive project outcomes, including data, in a cost-effective and feasible manner, similar to how projects were preserved 20 years ago for reuse in future projects. Additionally, the constant and exponential growth in the quantity of data created and stored in computer systems needs to be pinpointed. The two most voted driving factors towards cloud computing, availability, disaster recovery, and data backup, all contribute to the preceding scenario.

Security issues were selected as the primary impediment to utilizing cloud systems. All other selections received almost the same number of votes. This might entail a variety of things. The free-form responses provided valuable input on this subject, including the need for more broad guidelines on data management from legal, information technology, and cyber security. Education, communication, and standard tooling guidelines might make it easier for the majority to start using the cloud. If a guidebook outlining the safe ways to use the cloud were created, there might be a greater interest in using unfamiliar technology.

The author was particularly interested in the low amount of CAPEX/OPEX option votes; it is contingent on whether you are consuming or delivering services or budgeting and paying the bills. Financial stakeholders also have an important impact on the topic.

According to the responses to the questions "Who should be developing and operating cloud platforms," VTT IT is the most often chosen option, which corresponds to the free-form comments. VTT IT is a trusted service provider, and as such, a high level of service quality is demanded.

The survey's edge section received just a few replies due to the survey's intended target group not being reached. Nonetheless, the free-form feedback emphasized the critical nature of the open-source and the fact that the edge is a research topic in and of itself. Simply stated, "Hands-on solution development in comparison to commercial alternatives is critical since the edge represents a new paradigm." According to publicly accessible references, client market demand is increasing as businesses discover how to use the benefits of edge platforms to increase their profitability.

As with everyday things in life, there are the Pro-change and Anti-change behaviors in technology advance too - supporters, confused, passive, and active resistors to change. The most innovative and cutting-edge outcomes emerge from the "researcher's freedom" to break the rules. That is, typically adhere to regulations, but experiment with unconventional approaches as well. These are the factors that will drive research and innovation ahead.

Methods for advancing such innovation via improved service include promoting open-source, establishing guidelines, and packaging specific building blocks or environments to get started quickly. Clarifying unclear governance policies, effectively engaging with and training personnel on subjects that matter. Assuring cost transparency, having a clear and communicated solution for data sovereignty and lifecycle management, and, maybe most importantly, integrating security into the mix.

In summary, the findings indicate that a considerable culture transformation is required in which continuous improvement, current technologies, and proactiveness are ingrained in daily work to support research better.

The findings, conclusions, and a plan for internal development were produced and provided to IT management for further decision support.

4 SUMMARY AND CONCLUSIONS

The primary objective of this thesis was to compare cloud and edge platforms and to evaluate the scope of Microsoft Azure's present offering. Additionally, a survey was conducted at case study company VTT to assess the current state of cloud and edge platforms.

Latency is the primary disadvantage with cloud computing; while it is efficient at storing and processing data, it cannot help the distance between user and data center. High latency may become a serious concern as the number of edge devices rises dramatically.

Edge computing was created in reaction to the increasing proliferation of IoT devices, which are regularly linked to the Internet to receive data from the cloud or send data to it. Edge computing attempts to offer location awareness, reduce latency, and enhance service quality in real-time applications such as transportation, industrial automation, and actuator and sensor networks. Additionally, Edge computing enables real-time data analysis and helps in avoiding cloud security threats by adding computing capabilities to local devices near the users.

The term "edge" refers to any networking or computing resource located between cloud data centers and data sources. The edge itself must be well-designed in order to effectively satisfy network-wide requirements for features such as privacy, security, and reliability.

To summarize the differences between cloud and edge, the cloud offers lots of processing power, storage capacity, and large-scale analysis for non-real-time workloads. Bringing computing closer to the edge reduces the possibility of latency issues as well as network traffic to the cloud. Security will be the crucial component in the mix. Additionally, data sovereignty, contractual obligations, and legal compliance relating to data need to be emphasized.

As is true of any paradigm, edge computing presents several obstacles. Thus, the edge computing tale is far from over. Indeed, it has not properly started yet. While the author has sought to describe the future relevance of edge computing, it must be noted that this is a continuous process as new features and concerns emerge constantly. Nevertheless, the cloud needs to be brought to the edge via a distributed computing paradigm to get the best out of both worlds.

For the author, a key lesson point was discovering how difficult and how much expertise and skill it takes to conduct an effective survey that the target audience would find important. Due to the scarcity of responses, there is an opportunity for further development on the subject. Regardless, the conversations with the researchers gave a lot of insights and ideas where we are on the data revolution road and how IT can support by paving the road ahead.

In essence, it is becoming increasingly fascinating to see how technological advancements such as edge, cloud, distributed cloud, and the metaverse looming in the future may and will affect our lives. The author hopes that this thesis will generate more in-depth debates. This is not the finish line but rather the start.

REFERENCES

Ahmed, E., Ahmed, A., Yaqoob, I., Shuja, J., Gani, A., Imran, M. and Shoaib, M. (2017). Bringing Computation Closer toward the User Network: Is Edge Computing the Solution? *IEEE Communications Magazine*, 55(11), pp.138–144.

Alalawi, A. and Al-Omary, A. (2020). Cloud Computing Resources: Survey of Advantage, Disadvantages and Pricing. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/document/9325645 [Accessed 12 Apr. 2021].

Alliance for Internet of Things Innovation (2020). *Strategic Foresight Through Digital Leadership*. [online] . Available at: https://aioti.eu/wp-content/uploads/2020/10/IoT-and-Edge-Computing-Published.pdf [Accessed 7 Aug. 2021].

Ansari, M.S., Alsamhi, S.H., Qiao, Y., Ye, Y. and Lee, B. (2020). Security of Distributed Intelligence in Edge Computing: Threats and Countermeasures. In: T. Lynn, J.G. Mooney, B. Lee and P.T. Endo, eds., *The Cloud-to-Thing Continuum*. [online] pp.95–122. Available at: https://doi.org/10.1007/978-3-030-41110-7_6 [Accessed 28 Aug. 2021].

Bansal, N. (2020). *Designing Internet of Things Solutions with Microsoft Azure*. Berkeley, CA: Apress.

Bhamare, D., Zolanvari, M., Erbad, A., Jain, R., Khan, K. and Meskin, N. (2020). Cybersecurity for industrial control systems: A survey. *Computers & Security*, 89(), p.101677.

Cagnin, C., Muench, S., Scapolo, L., Stoermer, E. and Vesnic Alujevic, F. (2021). *JRC SCIENCE FOR POLICY REPORT Shaping & Securing THE EU'S OPEN STRATEGIC AUTONOMY by 2040 and beyond*. [online] Luxembourg: Publications Office of the European Union. Available at: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC125994/open_strategic_autonomy_2040_online.pdf [Accessed 9 Sep. 2021].

Calles, M.A. (2020). *Serverless Security*. Berkeley, CA: Apress.

Cao, J., Zhang, Q. and Shi, W. (2018). *Edge Computing: A Primer*. *SpringerBriefs in Computer Science*. Cham: Springer International Publishing.

Cisco (2021). *Secure Cloud for Azure (IaaS) Design Guide Design Guide Cisco public*. [online] . Available at: https://www.cisco.com/c/dam/en/us/solutions/collateral/enterprise/design-zone-security/scloud-azure-design-guide.pdf [Accessed 6 Aug. 2021].

Copeland, M., Soh, J., Puca, A., Manning, M. and Gollob, D. (2015). *Microsoft Azure: Planning, Deploying, and Managing Your Data Center in the Cloud*. Berkeley, Ca Apress.

Dhirani, L.L., Newe, T., Lewis, E. and Nizamani, S. (2017). *Cloud computing and Internet of Things fusion: Cost issues*. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/abstract/document/8304426 [Accessed 6 Aug. 2021].

Divvycloud (2020). *2020 State of Enterprise Cloud and Container Adoption and Security*. [online] Divvycloud.com. Available at: https://divvycloud.com/wp-content/uploads/2020/04/2020-State-of-Enterprise-Cloud-and-Container-Adoption-and-Security.pdf [Accessed 6 Aug. 2021].

Eclipse Foundation (2021). *IoT & Edge Commercial Adoption Survey 2021 Results*. [online] Eclipse Foundation. Available at: https://outreach.eclipse.foundation/cs/c/?cta_guid=e2521940-1aea-49d3-9411-0c2f1e707f6c. [Accessed 6 Aug. 2021].

European Union Agency for Cybersecurity (2019). *GOOD PRACTICES FOR SECURITY OF IOT Secure Software Development Lifecycle*. [online] . Available at: https://www.enisa.europa.eu/publications/good-practices-for-security-of-iot-1/at_download/fullReport [Accessed 6 Aug. 2021].

Gartner, Inc. (2018). What Edge Computing Means for Infrastructure and Operations Leaders. [online] www.gartner.com. Available at: https://www.gartner.com/smarterwithgartner/what-edge-computing-means-for-infrastructure-and-operations-leaders [Accessed 15 Aug. 2021].

Gartner, Inc. (2019). *Is The Cloud Secure*. [online] Gartner. Available at: https://www.gartner.com/smarterwithgartner/is-the-cloud-secure [Accessed 5 Aug. 2021].

Gartner, Inc. (2020a). 4 Trends Impacting Cloud Adoption in 2020. [online] Gartner. Available at: https://www.gartner.com/smarterwithgartner/4-trends-impacting-cloud-adoption-in-2020 [Accessed 3 Aug. 2021].

Gartner, Inc. (2020b). *The CIO's Guide to Distributed Cloud*. [online] Available at: https://www.gartner.com/smarterwithgartner/the-cios-guide-to-distributed-cloud/. [Accessed 19 Aug. 2021].

Gartner, Inc. (2021a). *Definition of Edge Computing - Gartner Information Technology Glossary*. [online] Gartner. Available at: https://www.gartner.com/en/information-technology/glossary/edge-computing [Accessed 11 Aug. 2021].

Gartner, Inc. (2021b). *Gartner Predicts the Future of Cloud and Edge Infrastructure*. [online] www.gartner.com. Available at: https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-cloud-and-edge-infrastructure/ [Accessed 8 Aug. 2021].

Gartner, Inc. (2021c). *Gartner Says Four Trends Are Shaping the Future of Public Cloud*. [online] Gartner. Available at: https://www.gartner.com/en/newsroom/press-releases/2021-08-02-gartner-says-four-trends-are-shaping-the-future-of-public-cloud. [Accessed 18 Aug. 2021].

Gartner, Inc. (2021d). *Gartner Predicts By 2025 Cyber Attackers Will Have Weaponized Operational Technology Environments to Successfully Harm or Kill Humans*. [online] Available at: https://www.gartner.com/en/newsroom/press-releases/2021-07-21-gartner-predicts-by-2025-cyber-attackers-will-have-we [Accessed 5 Aug. 2021].

Gartner, Inc. (2021e). Gartner Says Worldwide IaaS Public Cloud Services Market Grew 40.7% in 2020. [online] Gartner. Available at: https://www.gartner.com/en/newsroom/press-releases/2021-06-28-gartner-says-worldwide-iaas-public-cloud-services-market-grew-40-7-percent-in-2020 [Accessed 11 Nov. 2021].

Gartner, Inc. (2022). Gartner Predicts Hyperscalers' Carbon Emissions Will Drive Cloud Purchase Decisions by 2025. [online] Gartner. Available at: https://www.gartner.com/en/newsroom/press-releases/2022-01-24-gartner-predicts-hyperscalers-carbon-emissions-will-drive-cloud-purchase-decsions-by-2025 [Accessed 28 Jan. 2022].

Giannoutakis, K.M., Spanopoulos-Karalexidis, M., Filelis Papadopoulos, C.K. and Tzovaras, D. (2020). Next generation cloud architectures. In: T. Lynn, J.G. Mooney, B. Lee and P.T. Endo, eds. [online] Springer International Publishing, pp.23–39. Available at: https://doi.org/10.1007/978-3-030-41110-7_2. [Accessed 5 Aug. 2021].

Hajibaba, M. and Gorgin, S. (2014). A Review on Modern Distributed Computing Paradigms: Cloud Computing, Jungle Computing and Fog Computing. *Journal of Computing and Information Technology*, 22(2), p.69.

Hypothesis Group and Microsoft (2021). *IoT Signals EDITION 3 October 2021*. [online] Available at: https://azure.microsoft.com/mediahandler/files/resourcefiles/iot-signals/IoT%20Signals_Edition%202_English.pdf [Accessed 31 Oct. 2021].

IDG Communications, Inc. (2021). *What is edge computing and how it's changing the network*. [online] Network World. Available at: https://www.networkworld.com/article/3224893/what-is-edge-computing-and-how-it-s-changing-the-network.html [Accessed 12 Aug. 2021].

IEEE (2021). Real-life Use Cases for Edge Computing. [Online Image] Ieee.org. Available at: https://innovationatwork.ieee.org/wp-content/uploads/2019/06/Real-Life-Use-Cases-for-Edge-Computing_1024X684.png [Accessed 19 Nov. 2021].

Ifrah, S. (2020). Getting Started with Containers in Azure. Berkeley, CA: Apress.

Ikink, R. (2021). *Top cloud trends for 2021 and beyond*. [online] WordPressBlog. Available at: https://www.accenture.com/nl-en/blogs/insights/cloud-trends [Accessed 4 Aug. 2021].

Industrial Internet Consortium (2016). *Industrial Internet of Things Volume G4: Security Framework*. [online] . Available at: https://www.iiconsortium.org/pdf/IIC_PUB_G4_V1.00_PB.pdf [Accessed 6 Aug. 2021].

Industrial Internet Consortium (2018). Introduction to Edge Computing in IIoT. [online] Available at: https://www.iiconsortium.org/pdf/Introduction_to_Edge_Computing_in_IIoT_2018-06-18.pdf [Accessed 23 Apr. 2021].

Industrial Internet Consortium (2019). The Industrial Internet of Things Volume G1: Reference Architecture. [online] . Available at: https://www.iiconsortium.org/pdf/IIRA-v1.9.pdf. [Accessed 6 Aug. 2021].

Intel Corporation. (2021). Storage as a Service (STaaS). [online] Intel. Available at: https://www.intel.com/content/www/us/en/cloud-computing/storage-as-a-service.html [Accessed 12 Jan. 2022].

Iorga, M., Feldman, L., Barton, R., Martin, M.J., Goren, N. and Mahmoudi, C. (2018). Fog computing conceptual model. [online] Available at: https://doi.org/10.6028/NIST.SP.500-325 [Accessed 27 Nov. 2021].

Jensen, D. (2019). *Beginning Azure IoT Edge Computing*. Berkeley, CA: Apress.

Klein, S. (2017). *IoT Solutions in Microsoft's Azure IoT Suite*. [online] Berkeley, CA: Apress. Available at: https://link.springer.com/book/10.1007%2F978-1-4842-2143-3 [Accessed 6 Aug. 2021].

Kowalkowski, C., Gebauer, H., Kamp, B. and Parry, G. (2017). Servitization and deservitization: Overview, concepts, and definitions. *Industrial Marketing Management*, 60, pp.4–10.

Kubernetes (2021a). *Production-Grade Container Orchestration*. [online] Kubernetes. Available at: https://Kubernetes.io [Accessed 28 Sep. 2021].

Kubernetes (2021b). What Is Kubernetes. [online] Kubernetes.io. Available at: https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/ [Accessed 29 Sep. 2021].

Lacy, P., Daugherty, P., Durg, K., Ponomarev, P., Amin, K., Singh, S., Podder, S. and Sharma, S. (2020). *The Green Behind the Cloud | Accenture*. [online] www.accenture.com. Available at: https://www.accenture.com/nl-en/insights/strategy/green-behind-cloud [Accessed 4 Aug. 2021].

Lea, P. (2018). *Internet of Things for Architects*. S.L. Packt Publishing.

Lee, J., Leonardo, G., Milgram, J. and Rendón.D. (2021). *Azure strategy and implementation guide : up-to-date information for organizations new to Azure*. Birmingham: Packt Publishing.

Liu, F., Tang, G., Li, Y., Cai, Z., Zhang, X. and Zhou, T. (2019). A Survey on Edge Computing Systems and Tools. *Proceedings of the IEEE*, 107(8), pp.1537–1562.

Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J. and Leaf, D. (2011). NIST Cloud Computing Reference Architecture Recommendations of the National Institute of Standards and Technology Special Publication 500-292. [online] Available at: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication500-292.pdf. [Accessed 4 Aug. 2021].

Lynn, T., Endo, P.T., Maria, A., Barbosa, Gibson B. N and Rosati, P. (2020). The internet of things: Definitions, key concepts, and reference architectures. In: T. Lynn, J.G. Mooney, B. Lee and P.T. Endo, eds. [online] Springer International Publishing, pp.1–22. Available at: https://doi.org/10.1007/978-3-030-41110-7_1. [Accessed 21 Aug. 2021].

Machiraju, S. and Modi, R. (2018). *Developing Bots with Microsoft Bots Framework*. Berkeley, CA: Apress.

Mell, P. and Grance, T. (2011). *Special Publication 800-145 The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology*. [online] . Available at: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf [Accessed 4 Aug. 2021].

Microsoft (2020a). *Cloud anywhere: Azure for Hybrid and Multicloud Environments*. [online] azure.microsoft.com. Available at: https://azure.microsoft.com/en-in/resources/making-the-most-of-the-cloud-everywhere/ [Accessed 6 Aug. 2021].

Microsoft (2020b). *Download The Carbon Benefits of Cloud Computing: a Study of the Microsoft Cloud from Official Microsoft Download Center*. [online] www.microsoft.com.

Available at: https://www.microsoft.com/en-us/download/confirmation.aspx?id=56950 [Accessed 6 Aug. 2021].

Microsoft (2021a). *Azure Arc - Hybrid & Multicloud Management | Microsoft Azure*. [online] Microsoft.com. Available at: https://azure.microsoft.com/en-us/services/azure-arc/#product-overview [Accessed 17 Aug. 2021].

Microsoft (2021b). *Microsoft Azure IoT Reference Architecture*. [online] Available at: https://azure.microsoft.com/mediahandler/files/resourcefiles/microsoft-azure-iot-reference-architecture/Microsoft_Azure_IoT_Reference_Architecture_2_1_1_update.pdf [Accessed 20 Nov. 2021].

Microsoft (2021c). *Microsoft Cybersecurity Reference Architectures - Security documentation*. [online] docs.microsoft.com. Available at: https://docs.microsoft.com/en-us/security/cybersecurity-reference-architecture/mcra [Accessed 6 Aug. 2021].

Microsoft (2021d). *The twin transition: a new digital and sustainability framework for the public sector*. [online] Available at: https://wwps.microsoft.com/wp-content/uploads/2021/05/MSFT_EY-digital-sustainability-paper_final.pdf [Accessed 9 Nov. 2021].

Microsoft (2021e). *Zero Trust Cybersecurity for the Internet of Things*. [online] Available at: https://azure.microsoft.com/mediahandler/files/resourcefiles/zero-trust-cybersecurity-for-the-internet-of-things/Zero%20Trust%20Security%20Whitepaper_4.30_3pm.pdf [Accessed 20 Nov. 2021].

Microsoft (2021f). *Zero Trust Essentials eBook Zero Trust principles*. [online] Available at: https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RWIrfk [Accessed 20 Nov. 2021].

Microsoft (2021g). Azure Functions – Serverless Apps and Computing | Microsoft Azure. [online] azure.microsoft.com. Available at: https://azure.microsoft.com/en-us/services/functions/ [Accessed 14 Nov. 2021].

Microsoft (2022a). Zero Trust for the Microsoft identity platform developer. [online] Available at: https://azure.microsoft.com/mediahandler/files/resourcefiles/zero-trust-for-the-microsoft-identity-platform-developer/Zero%20Trust%20for%20the%20Microsoft%20identity%20platform%20developer.pdf [Accessed 9 Feb. 2022].

Microsoft (2022b). *Azure IoT – Internet of Things Platform | Microsoft Azure. azure.microsoft.com*. Available at: https://azure.microsoft.com/en-us/overview/iot/#industries. [Accessed 9 Jan. 2022].

Microsoft (2022c). Pricing—IoT Hub | Microsoft Azure. [online] azure.microsoft.com. Available at: https://azure.microsoft.com/en-us/pricing/details/iot-hub/ [Accessed 4 Jan. 2022].

Mijuskovic, A., Chiumento, A., Bemthuis, R., Aldea, A. and Havinga, P. (2021). Resource Management Techniques for Cloud/Fog and Edge Computing: An Evaluation Framework and Classification. *Sensors*, 21(5), p.1832.

Narayanan, A., Sena, A.S.D., Gutierrez-Rojas, D., Melgarejo, D.C., Hussain, H.M., Ullah, M., Bayhan, S. and Nardelli, P.H.J. (2020). Key Advances in Pervasive Edge Computing for Industrial Internet of Things in 5G and Beyond. *IEEE Access*, 8, pp.206734–206754.

Offin, N. (2020). Cloud computing vs. edge computing. [online] TechRadar. Available at: https://www.techradar.com/news/cloud-computing-vs-edge-computing [Accessed 4 Aug. 2021].

Orrin, S. and Chehreh, C. (2020). *How Edge Computing and Hybrid Cloud Are Shifting the IT Paradigm*. [online] Nextgov.com. Available at: https://www.nextgov.com/ideas/2020/11/how-edge-computing-and-hybrid-cloud-are-shifting-it-paradigm/170238/ [Accessed 5 Aug. 2021].

Overby, S. (2021). *8 edge computing trends to watch in 2021*. [online] IoT Central. Available at: https://www.iotcentral.io/blog/8-edge-computing-trends-to-watch-in-2021 [Accessed 11 Aug. 2021].

Pan, J. and McElhannon, J. (2018). Future Edge Cloud and Edge Computing for Internet of Things Applications. *IEEE Internet of Things Journal*, 5(1), pp.439–449.

Premsankar, G., Di Francesco, M. and Taleb, T. (2018). Edge Computing for the Internet of Things: A Case Study. *IEEE Internet of Things Journal*, 5(2), pp.1275–1284.

Red Hat, Inc. (2022). Edge computing brings data and insight closer to customers. [online] www.redhat.com. Available at: https://www.redhat.com/en/resources/bring-insight-data-customer-edge-computing-whitepaper [Accessed 18 Feb. 2022].

Ruth, J.-P.S. (2020). Exploring Edge Computing as a Complement to the Cloud. [online] InformationWeek. Available at: https://www.informationweek.com/infrastructure-as-a-service/exploring-edge-computing-as-a-complement-to-the-cloud [Accessed 23 Jan. 2021].

Sahai, A. (2020). *Council Post: To Be Secure, Enterprises Need To Really Understand The Cloud's Shared Responsibility Model*. [online] Forbes. Available at: https://www.forbes.com/sites/forbestechcouncil/2020/08/18/to-be-secure-enterprises-need-to-really-understand-the-clouds-shared-responsibility-model/ [Accessed 4 Aug. 2021].

Sahai, A. (2021). *Council Post: Distributed Cloud Is The Way Of The Future – What This Means For Your Business*. [online] Forbes. Available at: https://www.forbes.com/sites/forbestechcouncil/2021/06/21/distributed-cloud-is-the-way-of-the-future--what-this-means-for-your-business/ [Accessed 4 Aug. 2021].

Satyanarayanan, M. (2017). The Emergence of Edge Computing. *Computer*, 50(1), pp.30–39.

Section (2021). *Solving the Edge Puzzle*. [online] Available at: https://www.section.io/downloads/content-library/solving-the-edge-puzzle-white-paper.pdf [Accessed 5 Oct. 2021].

Serpanos, D. and Wolf, M. (2018). *Internet-of-Things (IoT) Systems*. Cham: Springer International Publishing.

Shi, J., Jin, L. and Li, J. (2019). The Integration of Azure Sphere and Azure Cloud Services for Internet of Things. *Applied Sciences*, 9(13), p.2746.

Shi, W., Cao, J., Zhang, Q., Li, Y. and Xu, L. (2016). Edge Computing: Vision and Challenges. *IEEE Internet of Things Journal*, 3(5), pp.637–646.

Shinder, D. (2019). *Trusted Cloud: Microsoft Azure Security, Privacy, Compliance, Reliability/Resiliency, and Intellectual Property*. [online] Microsoft.com. Available at: https://download.microsoft.com/download/1/6/0/160216AA-8445-480B-B60F-

5C8EC8067FCA/WindowsAzure-SecurityPrivacyCompliance.pdf [Accessed 6 Aug. 2021].

Shiyal, B. (2021). Beginning Azure Synapse Analytics. Berkeley, CA: Apress.

Soh, J., Copeland, M., Puca, A. and Harris, M. (2020). *Microsoft Azure*. Berkeley, CA: Apress.

Stackowiak, R. (2019). *Azure Internet of Things Revealed*. Berkeley, CA: Apress.

Svorobej, S., Bendechache, M., Griesinger, F. and Domaschka, J. (2020). Orchestration from the Cloud to the Edge. In: T. Lynn, J.G. Mooney, B. Lee and P.T. Endo, eds. [online] Springer International Publishing, pp.61–77. Available at: https://doi.org/10.1007/978-3-030-41110-7_4 [Accessed 8 Aug. 2021].

The Linux Foundation (2021). State of the Edge Report 2021. [online] State of the Edge. Available at: https://stateoftheedge.com/reports/state-of-the-edge-report-2021/ [Accessed 23 Nov. 2021].

Tsidulko, J. (2020). The 10 Biggest Cloud Outages Of 2020. [online] CRN. Available at: https://www.crn.com/slide-shows/cloud/the-10-biggest-cloud-outages-of-2020/2 [Accessed 17 May 2021].

Uddin, M.Y. and Ahmad, S. (2020). A Review on Edge to Cloud: Paradigm Shift from Large Data Centers to Small Centers of Data Everywhere. *2020 International Conference on Inventive Computation Technologies (ICICT)*.

Ullah, M., Nardelli, P.H.J., Wolff, A. and Smolander, K. (2020). Twenty-One Key Factors to Choose an IoT Platform: Theoretical Framework and Its Applications. *IEEE Internet of Things Journal*, 7(10), pp.10111–10119.

Ullah, M., Narayanan, A., Wolff, A. and Nardelli, P. (2021). Smart Grid Information Processes Using IoT and Big Data with Cloud and Edge Computing. [online] Available at: https://www.techrxiv.org/articles/preprint/Smart_Grid_Information_Processes_Using_IoT_and_Big_Data_with_Cloud_and_Edge_Computing/16995025 [Accessed 30 Nov. 2021].

VTT Technical Research Center of Finland Ltd. (2021). *VTT as a company | VTT*. [online] www.vttresearch.com. Available at: https://www.vttresearch.com/en/vtt-company [Accessed 22 Aug. 2021].

Wang, Y. and Leblanc, D. (2016). *Integrating SaaS and SaaP with Dew Computing*. [online] IEEE Xplore. Available at: https://ieeexplore.ieee.org/abstract/document/7723746 [Accessed 9 Mar. 2021].

Wortley, F. and Thompson, C. (2021). *Log4Shell: RCE 0-day exploit found in log4j2, a popular Java logging package | LunaSec*. [online] www.lunasec.io. Available at: https://www.lunasec.io/docs/blog/log4j-zero-day/ [Accessed 12 Dec. 2021].

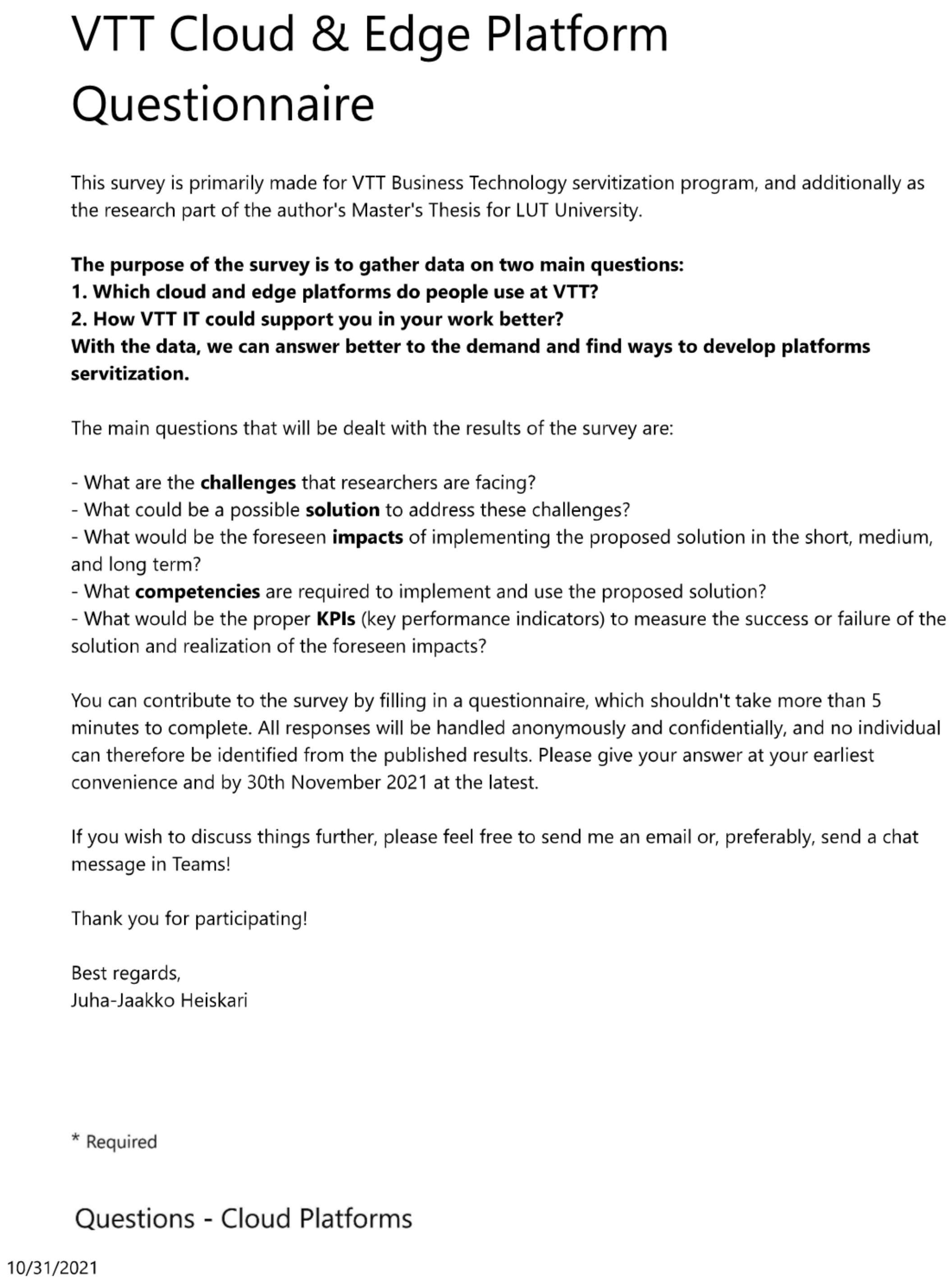
Wu, H., Zhang, Z., Guan, C., Wolter, K. and Xu, M. (2020). Collaborate Edge and Cloud Computing with Distributed Deep Learning for Smart City Internet of Things. *IEEE Internet of Things Journal*, 7(9), pp.1–1.

Xiao, Y., Jia, Y., Liu, C., Cheng, X., Yu, J. and Lv, W. (2019). Edge Computing Security: State of the Art and Challenges. *Proceedings of the IEEE*, 107(8), pp.1608–1631.

Yousefpour, A., Fung, C., Nguyen, T., Kadiyala, K., Jalali, F., Niakanlahiji, A., Kong, J. and Jue, J.P. (2019). All one needs to know about fog computing and related edge computing paradigms: A complete survey. Journal of Systems Architecture, [online] Volume 98. Available at: https://www.sciencedirect.com/science/article/pii/S1383762118306349 [Accessed 14 Nov. 2021].

Yu, W., Liang, F., He, X., Hatcher, W.G., Lu, C., Lin, J. and Yang, X. (2018). A Survey on the Edge Computing for the Internet of Things. IEEE Access, 6, pp.6900–6919.

Appendix 1. Microsoft Forms Questionnaire for VTT researchers



VTT Cloud & Edge Platform Questionnaire

This survey is primarily made for VTT Business Technology servitization program, and additionally as the research part of the author's Master's Thesis for LUT University.

The purpose of the survey is to gather data on two main questions:
1. Which cloud and edge platforms do people use at VTT?
2. How VTT IT could support you in your work better?
With the data, we can answer better to the demand and find ways to develop platforms servitization.

The main questions that will be dealt with the results of the survey are:

- What are the **challenges** that researchers are facing?
- What could be a possible **solution** to address these challenges?
- What would be the foreseen **impacts** of implementing the proposed solution in the short, medium, and long term?
- What **competencies** are required to implement and use the proposed solution?
- What would be the proper **KPIs** (key performance indicators) to measure the success or failure of the solution and realization of the foreseen impacts?

You can contribute to the survey by filling in a questionnaire, which shouldn't take more than 5 minutes to complete. All responses will be handled anonymously and confidentially, and no individual can therefore be identified from the published results. Please give your answer at your earliest convenience and by 30th November 2021 at the latest.

If you wish to discuss things further, please feel free to send me an email or, preferably, send a chat message in Teams!

Thank you for participating!

Best regards,
Juha-Jaakko Heiskari

* Required

Questions - Cloud Platforms

10/31/2021

1. What cloud platform(s) do you use? *

If your platform is not in any of the checkboxes, please select "Other" and write it there.

□ Microsoft Azure

□ Amazon Web Services

□ Google Cloud Platform

□ I don't use any cloud platforms

□ ____________________

Other

2. What type of cloud computing strategy do you consider good fit for your work? *

For examples of the "as a Service" please check wikipedia link: *https://en.wikipedia.org/wiki/As_a_service (https://en.wikipedia.org/wiki/As_a_service)*

□ Infrastructure as a Service (IaaS)

□ Platform as a Service (PaaS)

□ Software as a Service (SaaS)

□ Function as a Service (Faas)

□ Storage as a Service (StoaaS)

□ I don't think any of the strategies fit my work

□ ____________________

Other

3. What factors are driving your interest using cloud platforms? *

If you don't find a suitable option, please select "Other" and write it there.

□ Data Backup/Disaster recovery

□ Speed - quick to get workload running and quick to delete it

□ Always available access to applications and/or data

□ Cost savings

□ Scalability/flexibility

□ Improved productivity

□ Improved security options

□ Reduced workload

□ Transition from capital expenses to operating expenses

□ Customer market demand

□ None, I am not interested in cloud platforms

□

Other

4. What factors are obstacles for you using cloud platforms? *

If you don't find a suitable option, please select "Other" and write it there.

□ Security concerns

□ Integration with current environment

□ Availability

□ Legal and regulatory compliance

□ Transition from capital expenses to operating expenses

□ Company bureaucracy

□ Lack of knowledge/competence on deploying cloud services

□ None, I am not using cloud platforms

□ ____________________

Other

5. Who should be developing cloud technologies? *

Clarification: "Developing" means when you're testing/piloting a solution, and your workload is not yet in a production stage.

□ Myself

□ My research team

□ VTT IT

□ Partner

□ None of my concern

□ ____________________

Other

6. Who should be operating day-to-day cloud technologies? *

Clarification: "Operating" means when your workload is in a production stage, it's not test/pilot anymore.

□ Myself

□ My research team

□ VTT IT

□ Partner

□ None of my concern

□

Other

7. Is there anything else you would like to tell us about the adoption or use of cloud computing in your work?

Questions - Edge Platforms

8. What edge platform(s) do you use? *

If your platform is not in any of the checkboxes, please select "Other" and write it there.

□ AWS Greengrass

□ AWS Snowball

□ Azure IoT Edge

□ Bosch IoT Suite

□ Siemens MindSphere

□ I don't use any edge platforms

□

Other

9. What type of edge computing strategy do you consider good fit for your work? *

If you don't find a suitable option, please select "Other" and write it there.

□ Template/Starting kit of required components

□ Edge as a Service solution

□ I don't use any edge platforms

□

Other

10. What factors are driving your interest using edge platforms? *

If you don't find a suitable option, please select "Other" and write it there.

□ Artificial Intelligence

□ Containerization

□ Machine Learning

□ Source for Big Data analytics

□ Customer market demand

□ I don't use any edge platforms

□

Other

11. What factors are obstacles for you using edge platforms? *

If you don't find a suitable option, please select "Other" and write it there.

□ Security concerns

□ Integration with current environment

□ Availability

□ Legal and regulatory compliance

□ Transition from capital expenses to operating expenses

□ Company bureaucracy

□ Lack of knowledge/competence on deploying edge services

□ I don't use any edge platforms

□

Other

12. Who should be developing edge technologies? *

Clarification: "Developing" means when you're testing/piloting a solution, and your workload is not yet in a production stage.

□ Myself

□ My research team

□ VTT IT

□ Partner

□ None of my concern

□ ____________________

Other

13. Who should be operating day-to-day edge technologies? *

Clarification: "Operating" means when your workload is in a production stage, it's not test/pilot anymore.

□ Myself

□ My research team

□ VTT IT

□ Partner

□ None of my concern

□ ____________________

Other

14. Is there anything else you would like to tell us about the adoption or use of edge computing in your work?

How VTT IT could support you in your work better?

What are the IT oriented aspects which could be optimized for better performance from research perspective? What is the role of IT tools and support in enhancing the research process?

15. Please share your ideas how VTT IT could support you in your work better?

This content is neither created nor endorsed by Microsoft. The data you submit will be sent to the form owner.

Microsoft Forms

10/31/2021