UNIVERSITY OF TURKU

# Cloud versus Edge Computing in the Pharmaceutical Environment

A Decision-Making Framework for Regulated Environments

SMART SYSTEMS |

Department of Mechanical and Materials Engineering |

Master's thesis

Author:

Antton Laiho

Supervisors:

Professor Wallace Moreira Bessa

Sara Gambier

Jethro Lehtojärvi

09.10.2025

Turku

Master's thesis

**Subject**: Mechanical Engineering, Smart Systems
**Author**: Antton Laiho
**Title**: Cloud versus Edge Computing in the Pharmaceutical Environment
**Supervisor(s)**: Professor Wallace Moreira Bessa, Sara Gambier & Jethro Lehtojärvi
**Number of pages**: 92+11 pages
**Date**: 09.10.2025

As the pharmaceutical industry is adapting the principles from Industry 4.0 to Pharma 4.0 and at the same time researching the possibilities of new technologies such as artificial intelligence and machine learning in its highly regulated operating environment, the increasing amount of data that is collected from the manufacturing and business processes present many challenges. These challenges are not only coming from global and domestic regulation entities regarding e.g., computerized systems, GMP and data integrity. But also, from the usability of the data, regulatory adherence, and the technical challenges that the solutions itself have such as scalability, data privacy, latency and cost-efficiency All of these affect the decision-making regarding the computational solutions to be implemented.

This research aims to answer the research questions and to build decision-making framework to assist and standardise the decision-making process when Cloud and Edge computing solutions are evaluated and compared in highly regulated pharma environment.

**Key words**: AI, Cloud, Computing, Edge, Fog, GAMP, GMP, GxP, IoT, Industry 4.0, LLM, Pharma 4.0, Pharma

## Acknowledgements

I would like to thank all of the people participating in the surveys and interviews that were held during the time of writing and for those who gave valuable insights regarding the research topic. I also want to separately thank the supervisors for helping me through the whole process by giving me support when needed. Big thanks go also to the target company for which I conducted this research and to the innovation venture which enabled my participation in future life science development and to the consortium behind it. Both provided valuable knowledge, connections and topics for the thesis to be finished. Last but definitely not least, I want to show deep gratitude to my family, friends and supporters back home who enabled me to prioritize this work when other events overlap.

# Table of contents

**Acknowledgements** **4**

**1 Introduction** **11**

**2 Literature Review** **13**

**2.1 Regulated Operating Environment** **13**

2.1.1 EMA, FDA, FIMEA & ISPE 14

2.1.2 GMP, EU ANNEX 11 & GAMP® 16

2.1.3 Computerized systems 18

2.1.4 Industry 4.0, Pharma 4.0 and RAMI 4.0 19

**2.2 Cloud Computing** **26**

2.2.1 Definition and key concepts 27

2.2.2 Evolution and growth 31

2.2.3 Benefits and limitations 34

2.2.4 Architecture 37

**2.3 Edge Computing** **39**

2.3.1 Definition and key concepts 39

2.3.2 Evolution and growth 41

2.3.3 Benefits and limitations 43

2.3.4 Architecture 45

**2.4 Fog computing** **48**

**2.5 Large Language Models** **52**

**2.6 Literature summary** **54**

**3 Survey and Interview** **60**

**3.1 Survey design and focus group** **60**

**3.2 Results and conclusions** **62**

**4 Building guideline for decision making** **66**

**4.1 Contents of decision-making guideline and decision matrix** **66**

**4.2 Use-case examples evaluated with decision matrix** **70**

4.2.1 MES 70

4.2.2 Track and trace 71

4.2.3 Innovation venture 73

**5 Conclusions and discussion** **75**

**References** 78

**Appendices** 93

## List of figures

Figure 1. Pharma 4.0 framework (Adapted from *Pharma 4.0 Operating Model | ISPE, n.d.*). .............. 22
Figure 2. Value-based development stages (Adapted from *ISPE Pharma 4.0, 2023, p. 224*). ............ 23
Figure 3. RAMI 4.0 reference architecture (adapted from *Reference Architectural Model Industrie 4.0 | Plattform Industrie 4.0, 2018*). .......... 25
Figure 4. Cloud computing architecture (Adapted from *Shapla et al., 2020*). ...................................... 38
Figure 5. Cloud computing reference architecture by NIST (Adapted from *Liu et al., 2011, p. 3*). ....... 38
Figure 6. Edge architecture simplified (Adapted from *"What Is Edge Computing?," 2025*). ................. 46
Figure 7. Edge computing reference architecture (*Edge Computing Reference Architecture 2.0, 2017, p. 16*). .......... 47
Figure 8. Fog architecture simplified (Adapted from *Kazeem, 2018*). .................................................. 49
Figure 9. 3-Tier architecture of fog (Adapted from *Taneja & Davy, 2016*). .......................................... 50
Figure 10. Fog nodes and gateways (Adapted from *Sandoval, 2017*). ................................................. 50
Figure 11. Example transformer architecture (adapted from *Vaswani et al., 2017, p. 3*) ..................... 53
Figure 12. Cloud computing paradigm (adapted form *Shi et al., 2016, p. 638*). .................................. 56
Figure 13. Edge computing paradigm (adapted from *Shi et al., 2016, p. 638*). .................................... 57
Figure 14. Fog computing paradigm (Adapted from *Fog Computing and Internet of Things | Techplayon, 2017*). .......... 58
Figure 15. Whole computing stack (adapted from *Maciel et al., 2022, p. 235*). .................................... 58
Figure 16. Radar chart of question 1 answers. ...................................................................................... 63
Figure 17. Calculated values of question 1 answers. ............................................................................ 63
Figure 18. Radar chart of question 2-9 answers. ................................................................................... 64
Figure 19. Calculated values of question 2-9 answers. ......................................................................... 64
Figure 20. Enterprise spending on cloud datacentres including HW, SW and cloud infrastructure services (*Global Data Centre Spending | Statista, 2024*). .......... 94
Figure 21. Total size of public cloud market consisting of PaaS and IaaS (*Global Public Cloud Platforms Market | Statista, 2022*). .......... 94
Figure 22. Total size of global spending in public cloud services (*Public Cloud Computing Market Size | Statista, 2025*). .......... 95

## List of tables

Table 1. ALCOA++ principles and expectation of the data (*Adapted from ISPE Data Integrity by Design, 2020, pp. 11–12; EMA Guideline on Computerised Systems and Electronic Data in Clinical Trials, 2023, pp. 12–13*) ..... 24

Table 2. Three axes of RAMI 4.0 (*Reference Architectural Model Industrie 4.0* | *Plattform Industrie 4.0, 2018*)..... 26

Table 3. NIST cloud model (*Mell & Grance, 2011, 2-3*). ..... 27

Table 4. GAMP® 5 Responsibilities of supplier and customer (adapted from *ISPE GAMP® 5, 2022, 169*)..... 29

Table 5. Cloud benefits and their descriptions. ..... 34

Table 6. Cloud challenges and their descriptions according to (*Furht et al., 2010, pp. 17–18*)..... 36

Table 7. Edge benefits and their descriptions ..... 44

Table 8. Edge challenges and their descriptions ..... 44

Table 9. Characteristics of each paradigm (adapted from *Taheri & Deng, 2020, p. 10*). ..... 59

Table 10. Response rate and interviews held. ..... 61

Table 11. 1st version of the Decision matrix. ..... 67

Table 12. 2nd version of the Decision matrix. ..... 67

Table 13. 3rd version of the Decision matrix. ..... 68

Table 14. User defined importance – MES. ..... 71

Table 15. User defined importance – Track and Trace. ..... 72

Table 16. User defined importance – Innovation venture. ..... 73

## List of Abbreviations and Acronyms

| | |
|---|---|
| 4IR | Fourth industrial revolution |
| AI | Artificial intelligence |
| AIaaS | Artificial Intelligence as a Service |
| AII | Alliance of industrial Internet |
| ALCOA+(+) | Attributable, Legible, Contemporaneous, Original, Accurate, Consistent, Enduring and Available (and Traceable) |
| ARPANET | Advanced Research Projects Agency networks |
| AWS | Amazon Web Services |
| CaaS | Containers as a Service |
| CapEx | Capital Expenditure |
| cGMP | Current Good Manufacturing Practice |
| CoP | Community of Practice |
| CPU | Central Processing Unit |
| CS | Computerized systems |
| DevOps | "Development" and "Operations", A set of practices for automating the processes between software development and information technology operations teams so that they can build, test, and release software faster and more reliably. |
| ECC | Edge Computing Consortium |
| EEA | European Economic Area |
| EMA | European Medicines Agency |
| EMEA | European Medicines Evaluation Agency |
| ETSI | European Telecommunications Standards Institute |
| EU | European Union |
| EU Annex 11 | European Union’s GMP guideline - Annex 11 - Computerized systems |
| FaaS | Function as a Service |
| FD&C Act | Federal Food, Drug, and Cosmetic Act |
| FDA | U.S. Food and Drug Administration |
| Fimea | Finnish Medicines Agency |
| FinOPS | “Finance” and “DevOps”, stressing the communications and collaboration between business and engineering teams |
| FW | Firmware |
| GAMP® | Good Automated Manufacturing Practice |
| GDPR | General Data Protection Regulation |
| GPU | Graphics Processing Unit |
| GxP | Good x Practice – Collectively referring to GCP, GDP, GLP, GMP, GVP |
| HIPAA | Health Insurance Portability and Accountability Act |
| HW | Hardware |

| | |
|---|---|
| IaaS | Infrastructure as a Service |
| ICT | Information and Communication Technology |
| IoT | Internet of Things |
| 4IR | 4th Industrial Revolution |
| ISPE | The International Society for Pharmaceutical Engineering |
| IT | Information Technology |
| MEC | Multi-access Edge Computing (formerly known as Mobile Edge Computing) |
| ML | Machine learning |
| MMR | Mixed Methods Research |
| NCA | National Competent Authority |
| NIST | National Institute of Standards and Technology |
| NLP | Natural Language Processing |
| NLU | Natural Language Understanding |
| NN | Neural Network |
| OpEx | Operating Expense |
| OS | Operating System |
| OT | Operational Technology |
| PaaS | Platform as a Service |
| PIC/S | Pharmaceutical Inspection Convention and Pharmaceutical Inspection Cooperation Scheme |
| RAMI 4.0 | Reference Architectural Model for Industrie 4.0 |
| RT | Runtime |
| SaaS | Software as a Service |
| SLA | Service Level Agreement |
| SW | Software |
| XaaS | Anything “X” as a Service |

# 1 Introduction

This thesis utilizes mixed-method research based on literature review, survey and semi-structured interviews. Explanatory sequential design form is used to first collect quantitative data and analyse it and based this the qualitative data collection and analysis follows. This method and design form allows the use of both, quantitative and qualitative method and the strengths of each to better understand the research problems and test the hypothesis of decision-making framework in regulated pharma environment. Mixed-method research explicitly combines both deductive and inductive perspectives.

Artificial Intelligence (AI) or more specifically, Generative AI, has been used to help create few images in this thesis. These images have been marked with footnote to highlight that they have been created using OpenAI's Sora.

The goal of this thesis is to study and compare two computing paradigms and to answer the research questions presented below. This is done by first collecting and analysing the quantitative data and to enrich the results, qualitative approach is being applied.

Research questions:

1. How to determine when pharma regulated data is better to be processed in cloud over edge?
2. Is it possible (and /or needed) to create guideline that can be used during decision-making in regulated pharma environment?

Thesis starts with the literature review, which gives the ground for the research and basis for building survey questions and list most important characteristics that influence the decision-making. The survey is sent to hand-picked group of individuals related to the innovation venture, this thesis or the work packages that are done in the target company. These individuals have different level of technical background, and the aim is to get the characteristics collected from the literature revive to be rated. After the survey is done, each respondent is requested to join small semi-structured interview regarding the survey where they are able to give more broader answers than numeric answer on scale from 1 to 4. The purpose of the interview is to get more broader insight on the importance of individual characteristic of cloud and edge in regulated pharma environment, which quantitative research cannot provide by itself, and it gives deeper understanding on the results of the survey. Based on these three, the decision matrix is built and weighted, and the decision-making guideline is formed to support the use of the matrix.

Literature revive starts with the regulated pharma environment and the constrains it bring to the thesis. This section brings the knowledge of the reader to the level that they are able to understand the high regulation related to pharma environment and how it affects the results of this work. From there, the review proceeds to the paradigms that are being evaluated. Cloud computing and Edge computing are studied more thoroughly, and Fog computing is merely presented as a tertiary option and future research in mind. Last part of the review gets acquainted with Large Language Models (LLMs) and how they could be used in the decision-making process to make it faster and more repeatable.

Based on the quantitative data collected from the literature review, a survey is deductively formed and then conducted. The survey questions are built to question the relevancy of the characteristics that arose from the literature to the highly regulated working environment of pharma, and it aims to collect valuable information for the making of decision matrix which act as a standardised tool to make decisions regarding computation solutions. After respondents have answered the survey, a semi-structured interview is held to get more in-depth understanding of the answers each respondent has given and also their thoughts and opinions are listened regarding the research topic. This interview enriches the qualitative data with quantitative data and makes the thesis mixed-method research. The final decision matrix is then formed based on the combination of these. A decision-making guideline is formed around the decision matrix to instruct the use of the matrix and to give constraints and mandate to the person using the guideline.

To test the hypothesis that decision-making guideline can be used in the decision-making process in regulated pharma environment, three use-case examples are presented from the target company that are tested against the guideline and the decision matrix, and the results are evaluated. This shows how well the guideline, and the decision matrix have been designed and materialized, and if they work in the regulated environment with scalable use-cases.

In conclusions and discussion, the results of the research are gone through and the role of LLMs and Fog computing are discussed. Also, few development ideas and ideas for future research are brought up.

# 2 Literature Review

The aim of this literature review is to give reader the relevant knowledge of the regulated pharma environment which gives this thesis work its boundary conditions, next it will help the reader to understand the technical characteristics and backgrounds of the comparable technologies, Cloud computing and Edge computing, that are evaluated in this thesis and finally it goes through the theoretical part of building structured guideline for decision making. In addition to this, fog computation and large language models are briefly introduced. At the end, the reader is presented with a summary of the most significant takes of the literature review to give clear understanding of the theoretical background on which the research is built upon and to help reader to proceed to the next chapters which go deeper to the research trying to give answers to the research questions presented in the Chapter 1.

## 2.1 Regulated Operating Environment

In pharmaceutical industry the regulation is heavy and there are multiple different entities that are part of creating, maintaining and monitoring the laws, regulations, guidelines, and that they are followed by the pharmaceutical companies around the world. The regulation consists of different functions e.g., clinical trial evaluation, inspection and licensing of pharmaceutical manufacturing sites and distribution channels associated with them. These entities are for example European Commission, European Medicine Agency (EMA), U.S. Food and Drug Administration (FDA) and International Society for Pharmaceutical Engineering (ISPE). They work in collaboration to promote human health by streamlining effort, sharing best practices and to prevent duplication. The purpose of the regulation is to ensure patient safety, with safe and high-quality pharmaceuticals that are also effective with favourable risk-to-benefit ratio. (*Pharmaceutical Industry | Definition, Overview, History, Characteristics, Examples, & Facts | Britannica, 2025; United States | European Medicines Agency (EMA), 2023.*) In the following chapters, the regulatory agencies and regulations/practices which are important for the Finnish pharmaceutical industry and important for this particular thesis, are shortly introduced. The purpose for this introduction, is to help reader understand the regulatory burden associated with this research and to get basic knowledge of the pharmaceutical industry in general, pharmacovigilance that requires drug companies which have marketing authorisation to monitor the safety of medical products and Finnish Medicines Agency (Fimea) is the authority that oversees this in Finland (*Pharmacovigilance | Fimea, n.d.*) and

Good x Practices (GxP) compliance related to computerized systems in order to fully benefit of the research and results presented in the upcoming chapters.

### 2.1.1 EMA, FDA, FIMEA & ISPE

**The European Medicines Agency** (EMA) has been operating since 1995, to protect human and veterinary health in European Union (EU) member states and other countries in the European Economic Area (EEA). EMA was originally known as European Medicines Evaluation Agency (EMEA) and this name can be found in older literature. It is decentralised agency which responsibility is to protects the public and animal health by making sure that the medicines sold in the EU market are safe, effective and high quality. This is possible through supervision, scientific evaluation and safety monitoring of medicine. (*About Us | European Medicines Agency (EMA), 2016.*) EMA is closely working together with the European Commission and USA Food and Drug Administration to avoid duplication, share best practices and streamline effort in order to promote human and animal health and they have confidentiality arrangements to permit exchange of commercially confidential information regarding scientific and regulatory processes. These arrangements have been in place since 2003. These entities also have mutual recognition agreement that allows both parties, FDA and their counterparts EU authorities to rely on each other's Good Manufacturing Practise (GMP) inspections, Waive batch testing of products on entry into their territories and to share information on quality defects and inspections. (*United States | European Medicines Agency (EMA), 2023.*)

European union lists the beneficiaries of the work that EMA is doing, these are not limited only to humans and animals, but also e.g., pharmaceutical companies and academics are recognised to gain from the EMA guidelines, scientific advice programme and incentives, and other responsibilities that EMA has (*European Medicines Agency – EMA | European Union*, *2024*).

EMA develops technical guidance and gives scientific support and advice to national competent authorities such as Fimea in Finland, who have role in new medicine clinical trials regionally. Despite the centralized authorisation process, which requires pharmaceutical companies developing new innovations to submit marketing-authorisation applications to EMA, EMA does not permit marketing rights to medicinal products in the EU, it is the European Commission who takes a legally binding decision regarding the permits and EMA only gives recommendations for this decision making and coordinates the processing of the applications. EMA ensures that the pharmaceutical companies follow standard that come from EU and other international bodies which benefits the company when European Commission assess whether to grant the marketing-authorisation or not. (*About Us | European Medicines Agency (EMA)*, *2016.*)

**U.S. Food and Drug Administration** (FDA) has been promoting and protecting public health almost 120 years, since 1906. FDA monitors all drugs that are manufactured and sold to the U.S. market regardless of the manufacturing country. The drug manufacturers need FDA approval before they can enter the U.S. market. This approval requires manufacturers to follow and comply the law regarding drug safety and quality, and relevant regulations of the Federal Food, Drug, and Cosmetic Act (FD&C Act) and Current Good Manufacturing Practice (cGMP). The approval process for the drug application consists of assessment of the faculty where the drug(s) are manufactured and the manufacturing process itself. (*About | FDA, 2024; "Pharmaceutical Inspections and Compliance | FDA," 2024.*)

**Finnish Medicines Agency** (Fimea) is the National Competent Authority (NCA) which is part of European medicine regulatory networks which include European commission, EMA and other NCAs. Operating under Ministry of Social Affairs and Health in Finland the central administrative agency promotes health and safety of people. This is done by regulating e.g., medical devices and medicinal products. The agency supervise that the manufacturers comply with the regulation and they oversee and carry out pharmacovigilance of medical products. (*About Us | Fimea, 2022.*)

Applying of the marketing authorization and registration of pharmaceutical products in Finland and in the whole European Union is possible with several alternative procedures. Suomi.fi Web service has guidance on the ways that the authorization can be applied. These are listed below. (*Marketing Authorisations | Suomi.Fi, 2025.*)

- National procedure – Can be applied when product does not have marketing authorization in another EU country, Norway or Iceland. The application is only submitted to Fimea who evaluates the application and decides whether to grant the authorization or not.

- Mutual Recognition Procedure – When product already has national authorization in other EU country, Norway or Iceland, Fimea grants the marketing authorization in Finland.

- Decentralized Procedure – Marketing authorization can be granted to multiple countries at the same time.
- Centralized Procedure – Marketing authorization is applied for all EU countries, Norway and Iceland at the same time. This application is submitted to EMA which coordinates the application process and the European Commission grants the authorization.

**International Society for Pharmaceutical Engineering** (ISPE) is a nonprofit association which aim is to lead technical, regulatory, and scientific advancements in throughout the entire lifecycle of pharmaceutics. They are connecting global industry leaders and their knowledge to deliver operational excellence, supply chain innovation, and regulatory insights to enable development, manufacturing and reliability of medicine. ISPE has been founded in North America in 1980 and today their membership base has grown to cover 120 countries and more than 21 000 pharmaceutical professionals, and they cover 20 unique technical areas such as Good Automated Manufacturing Practice **(**GAMP®) and PHARMA 4.0 with their ISPE Communities of Practice (CoP). They focus on regulation impacting facilities, manufacturing processes and operations, and supply chain sustainability over the product lifecycle. ISPE as a society leads and facilitates next-gen process technology and innovative technical solution development. (*About Us | ISPE, n.d.*.)

### 2.1.2 GMP, EU ANNEX 11 & GAMP®

**Good Manufacturing Practice** (GMP) regulations give the minimum requirements for drug manufacturing. It covers manufacturing facilities, methods and the controls that are used in the processing, packaging and manufacturing of the drug. In the U.S., these regulations are enforced by FDA and in Europe, the rules governing medicinal products is European Commissions guidelines and they are published under EudraLex publication as EudraLex – volume 4. The enforcement of these guidelines is done by National Regulatory Agencies, by EMA in Finland. (*EudraLex - European Commission*, 2025; *“Facts About the Current Good Manufacturing Practice (cGMP) | FDA,” 2025; Good Manufacturing Practice | European Medicines Agency (EMA), 2022.*)

GMP has long history dating back to 1960s when the world’s first legally binding GMP regulations got their legal basis in the U.S. From this, the regulation has evolved quite significantly, and it has spread across the world. First WHO published their GMP guideline and from there Europe followed. (*What Is GMP? - ECA Academy, n.d.*) Referring to the most recent GMP regulations letter "C" is added in front of the acronym to denote "current". This same naming method is applied to all the Good "x" Practices (GxP) acronyms/definitions. The CGMP requires manufacturers to use up-to-date systems to qualify with the "current" regulations. *(Acronyms and Abbreviations | FDA, n.d.; "Facts About the Current Good Manufacturing Practice (CGMP) | FDA," 2025.*)

**EU ANNEX 11** applies to computerized systems used in GMP- regulated activities (*"EudraLex Volume 4 - Annex 11," 2011*). It is guideline, not legally binding regulation, and regulatory authorities can consider manufacturers complying Annex 11 to meet data integrity and quality control standards. In the pharmaceutical industry, following Annex 11 is regarded as best practice. (*Caetano, 2023.*)

European commission lays down the principles and guidelines of GMP in EudraLex – Volume 4 which governs medicinal production in the EU. Annex 11 is part of the guidance for interpretation of the GMP in human and veterinary medicinal product manufacturing. (*EudraLex - Volume 4 - European Commission*, *2025.*)

Eu annex 11 gives guidance that covers computerized system risk management, validation, audit trails, change and configuration management, security, business continuity and archiving. It also guides on the use of suppliers and service providers. (*"EudraLex Volume 4 - Annex 11," 2011.*) Validation is one of the most critical aspects of the whole lifecycle of GPM relevant computerized system, "this not only ensures compliance of regulation, but it also ensures product quality and prevents possible costly product recalls and regulatory penalties – production without validated computerized systems is prohibited by the company processes and if such case happens when production has been done without the validated status then evaluation of the situation, deviation process and product quality risks need to be done." (*H. Soinio, personal communication, September 25, 2025*).

**Good Automated Manufacturing Practice** (GAMP®) is an ISPE Community of Practice (CoP), which is technical subcommittee under the International Society for Pharmaceutical Engineering (ISPE), and guidelines compiled for the pharmaceutical industry. These guidelines are meant to  safeguard product quality, patient safety, and data integrity in the context of GxP computerized systems. (*What Is GAMP® | ISPE, n.d.*)

GAMP® 5 Guide, 2nd edition is the newest publication (July 2022) of the GAMP®. It is a risk-based approach which seeks to exceed minimum expectations of the compliance by encouraging to apply modern, current, good Information Technology (IT) practices and excellence in software engineering. These enable higher product safety and quality from which the patient and the public benefit. The 2nd edition maintains 1st editions principles and framework. These include increased importance of evolving approaches in software development and increased use of automation and software tools. (*What You Need to Know About GAMP® 5 Guide, 2nd Edition | Pharmaceutical Engineering, 2025.*)

The 2nd edition has been revised, and it has many new topics for example, the new appendix D11 – Artificial Intelligence and Machine Learning (AI/ML) are added, and other relevant topics are updated, such as cloud computing. Pharmaceutical industry recognises the importance of AI and ML in future pharma environment and this guidance is risk based, and it covers the whole life cycle of the ML component or sub-system. AI/ML appendix suggests that successful implementation of ML requires e.g., thorough business analysis and process understanding, application of good software development and engineering practises. It also suggests that even tough traditional computerized-system lifecycle and compliance/validation approaches are completely applicable, verification and validation approaches should be designed to consider change and configuration management, version control and monitoring to cope with the differentiation of outputs as the ML sub-system evolves. (*ISPE GAMP® 5, 2022, pp. 10 & 269–270.*)

Important aspects and guidance that the GAMP 5 gives related to this thesis are the following:

- Cloud infrastructure and service models – Page 169-170 Appendix M11 (see also thesis Chapter 2.2.1)
- Cyber security – Page 331 Appendix O11
- Computerized systems validation – Page 21-27
- IT infrastructure qualification – Page 163-170 Appendix M11

### 2.1.3 Computerized systems

Computerized systems (CS) include, according to Pharmaceutical Inspection Convention and Pharmaceutical Inspection Cooperation Scheme (PIC/S), computer Hardware (HW) and software (SW), installed devices and controlling SW of the computer. The whole system is composed of both the computer system and function or process controlled by the CS. (*PIC/S Secretariat - PI 011-3, 2007, 7.*)

Annex 11 of EudraLex – Volume 4 GMP guidelines, is governing CS in all GMP regulated processes. In this annex, the use of electronic records and signatures are detailed alongside with the guidelines for data integrity, security and validation. The importance of Annex 11 for this research is quite substantial. It gives the requirements for the validation of computerized systems and qualification of the IT. Whether it is outsourced Cloud service or locally hosted on-premises Edge solution, Annex 11 (*2011*) applies to both ways of managing and storing data.

In GAMP® 5, guidance for GxP compliant CS is provided with high detail. GAMP provides guidance, that is practical for the industry, by establishing language and terminology that is common for the whole regulated environment, it promotes system lifecycle approach, facilitates regulatory requirements interpretation and finally it aims to clarify roles and responsibilities regarding risk-based approach to CS (*ISPE GAMP® 5, 2022, 9*). Regarding the overall safety and resilience of the computerized systems, cybersecurity risk management plays key role in the evaluation of threads in the GxP environment (ISPE GAMP® 5, 2022, 166). In GAMP® 5 (*2022*), the adoption of cloud-first strategies and "as a Service" solutions are acknowledged and these potential risks associated with the use of cloud are directly related to the research topic of this thesis. As the CS can be locally hosted in private servers or in "public" cloud that is provided by cloud solutions vendor e.g., Amazon Web Services (AWS), the relevancy of the GAMP 5 remains high.

### 2.1.4 Industry 4.0, Pharma 4.0 and RAMI 4.0

**Industry 4.0** also referred as 4th Industrial Revolution (4IR) or information age in the literature, is the "current" industrial revolution that begun in the 21th century (*Gordan et al., 2023, pp. 1–2*).

Gordan, Ghaedi and Saleh (*2023*) list the key innovations of 1-4IR and their impacts in clear and understandable manner. For the 4IR, these are as follows.

Key innovations:

- Cognitive computing

- Big Data, Cloud Computing, and Internet of Things (IoT)
- Machine-to Machine/Human Communication
- Cyber Physical Systems

Impacts:

- Data and information explosion
- Arrival of smart, virtual, and augmented systems
- Sustainable living, solutions and societies

4IR presents many benefits to the industry. According to multiple studies (*Sony et al., 2021; Brozzi et al., 2020; Kiel et al., 2017; Thames & Schaefer, 2017*), these benefits have significant impact to the businesses implementing Industry 4.0. Some of these benefits are as listed below.

- Improving customer satisfaction
- Maximize efficiency, cost cutting in operations and remain competitive
- Organizational decisions driven by big data
- Ability to answer increased smart products demand
- Organizational efficiency and agility

4IR holds in AI under the cognitive computing bullet point, and big data, cloud computing and IoT that are relevant to this research. Cognitive computing refers to a simulation of "human-like" thought process of AI, where machines are capable of learning, reasoning and understanding natural language (*Hurwitz et al., 2015; Gordan et al., 2023, p. 29; Cognitive Computing | Google, n.d.*). AI as a concept requires lots of processing power and data to produce applications and/or functions that create value to businesses, and these are vital aspects considered in this research. AI, IoT and big data go hand in hand as AI requires vast amounts of data, IoT produces lots of data and big data is extensive dataset in the characteristics of volume, velocity, variability and/or variety (*Gordan et al., 2023; Chang et al., 2018, pp. 8–9*). IoT is a network of HW, SW, firmware (FW) and actuators that are connected and accessible via internet (*Banafa, 2023; Glossary | NIST, n.d.*).

The most significant impact of 4IR comes from the data and information explosion. Also, arrival of smart, virtual, and augmented systems plays key role in this research due to the parameters that these impacts hold within, for example smart factories that connect cyber world to the real world via automation and monitoring that requires real time data processing and -controls (*Gordan et al., 2023, p. 12*). Data and information explosion comes from many sources e.g., streamed media and IoT telemetry data (*Gordan et al., 2023, p. 5*). This trend in which data is created today and the estimation in which it is going to grow in the future is exponential. This growth follows patterns that can be described by known information society laws e.g., Moore's law, Kryder's law and Bell's law. These describe the exponential growth in particular field and not directly the growth of data and information explosion. (*Gams & Kolenik, 2021.*)

According to Gordan et al. (*2023*), the five enabling technologies of 4IR are IoT & sensor networks, big data and data analytics, Information and Communication Technology (ICT) & networking, AI & ML, and cloud, fog & edge computing. From these, this thesis is focusing on the research between cloud, fog & edge computing and the contributing elements that might affect the result of the research questions. In the upcoming chapters, each of these cloud, fog & edge computing are studied individually.

The 4IR paradigm integrates new technologies and employs information created in different stages of the lifecycle of the product to better optimize and utilize used resources, and to make higher quality products with better performance (*Papakostas et al., 2020, p. 1*). Gajdzik and Wolniak (*2022*) did a study on the influence of 4IR to projects on business operations and they listed the most influencing categories to enterprises, top three being quality & personalisation, speed & agility and adaptability & work precision. This study was made of steel enterprises segment in Poland. They conclude that modifying production to implement 4IR and changing towards smart production, perceive benefits of e.g., higher product quality, and other Industry 4.0 benefits (*Gajdzik & Wolniak, 2022, p. 14*).

The future and evolution of 4IR is changing towards industry 5.0 which, according to European Commission (*2024*), is already undergoing. Transitioning from 4IR to 5IR not only benefits the industry, but also the workers and the society (*Industry 5.0 - European Commission, 2024*). It aims to integrate human intelligence and advanced technologies such as AI to enable sustainability and human centric development (*Aaltonen & Kurvinen, 2025;*

*Nousala et al., 2024*). This study will not go deeper into the future of Industry 4.0 or to Industry 5.0.

**Pharma 4.0™** resembles evolution where digitalization and automation meet complex product portfolios and cycles. The cornerstone of Pharma 4.0 is digitalization that connects "everything" in smart plant floor enabling fast decision-making and control over operations, business, regulatory compliance and quality (*Pharma 4.0™ | ISPE, 2024*). The concept of Pharma 4.0 evolves from the revolution of Industry 4.0 and compared to rest of the industry sectors; pharma is slowly adapting to the 4IR. Few reasons for this are the heavily regulated environment which causes  lag in the adoption and also the complexity of the manufacturing and validation are key factors. (*Buntz, 2021; Pharma 4.0 for plant manufacturing | Schneider Electric, n.d.*)

ISPE has built baseline guideline for life science organizations to help them start, restart, or continue their journey to achieve the goals in the 4IR and Pharma 4.0 transformation. Triggers for the transformation towards Pharma 4.0 are the challenges that industrial digital transformation brings and the opportunities it could create to the businesses and the possibility to create value for the customers. The guideline lists big changes that have impacted pharmaceutical industry in the recent years. These include advances in biotechnology, personalized medicine, and digitization & digitalization. Addition to these, increased globalization has also been big trigger for the pharma industry providing access to expanded markets and this results in reaching wired customer base. (*ISPE Pharma 4.0, 2023.*)

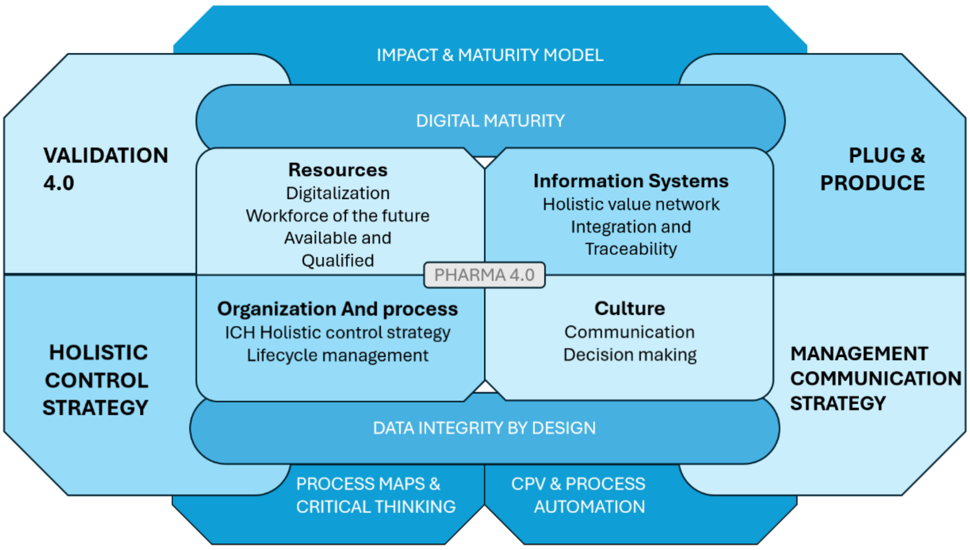



Figure 1. Pharma 4.0 framework (Adapted from *Pharma 4.0 Operating Model* | *ISPE, n.d.*).

Figure 1 shows framework for Pharma 4.0 operating model and the operating model CoP work groups and the prerequisites needed to enable successful implementation for business-driven digital transformation. This operating model is based on the 4IR concept, and it focuses on information systems, culture, organization and processes, and resources. The operating model is holistic strategy for controlling lifecycle management with data integrity (DI) by design based on the 4IR architectural framework of organization's digital maturity. This guideline provides tools for the assessment of the organization's digital maturity. (*ISPE Pharma 4.0, 2023.*)

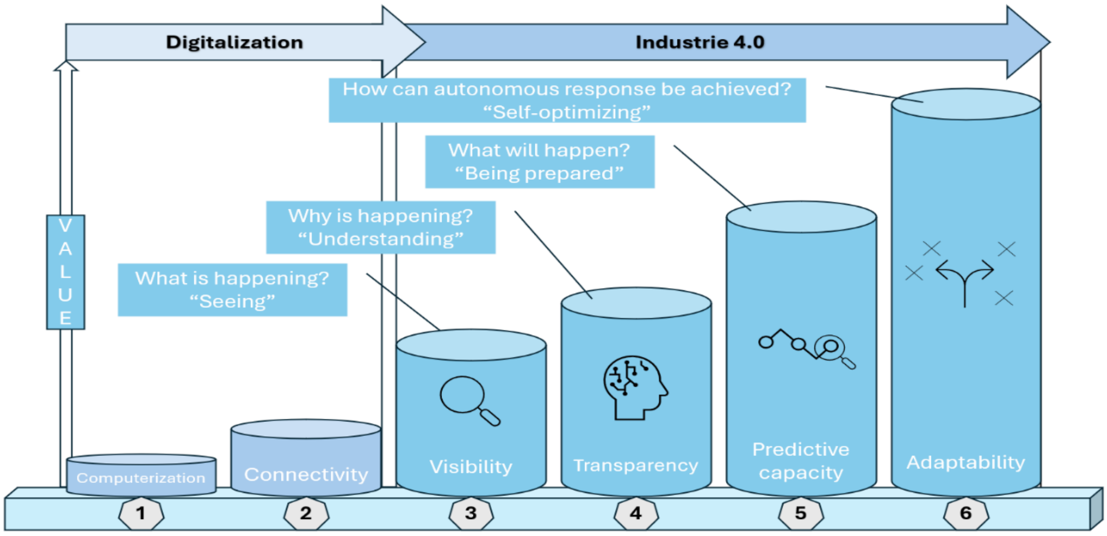



Figure 2. Value-based development stages (Adapted from *ISPE Pharma 4.0, 2023, p. 224*).

Figure 2 shows value-based development stages of the digital maturity model that can be used to assess the organization's level of maturity regarding the Pharma 4.0 operating model. Digital maturity is one of the two enablers for the Pharma 4.0 operating model, seen in Figure 1 serving as an enabler and prerequisite for the maturity self-assessment tool of Pharma 4.0. This self-assessment tool is designed to allow organizations to assess their Pharma 4.0 maturity and define needed improvement actions based on the results. (*ISPE Pharma 4.0, 2023, pp. 224–225.*)

Data integrity is an important factor and the second key enabler for the Pharma 4.0 (*ISPE Pharma 4.0, 2023, p. 72*). It is defined by National Institute of Standards and Technology (NIST) as "The property that data has not been altered in an unauthorized manner.", NIST also defines that "Data integrity covers data in storage, processing, and while in transition." (*Glossary | NIST, n.d.*). One of the key terms and principles of the ISPE Data Integrity by Design Good Practice Guide is the Attributable, Legible, Contemporaneous, Original, Accurate, Consistent, Enduring and Available (ALCOA+) -principles. In the Table 1 the principles and the data expectations associated with each is shown. This table also presents the new version of EMAs Guideline on computerised systems and electronic data in clinical trials which introduces one more character to the principle, traceability, making it ALCOA++.

Table 1. ALCOA++ principles and expectation of the data (*Adapted from ISPE Data Integrity by Design, 2020, pp. 11–12; EMA Guideline on Computerised Systems and Electronic Data in Clinical Trials, 2023, pp. 12–13*)

| **Principles** | **Data Expectation** |
|---|---|
| Attributable | Attributable to the person or system generating the data |
| | Identify the person or system performing an activity that creates or modifies data |
| | Linked to the source of the data |
| Legible | Readable and permanent |
| | Accessible throughout the data life cycle |
| | Original data and any subsequent modifications are not obscured |
| Contemporaneous | Recorded or observed at the time the activity is performed |
| Original | Original data is the first recording of data, or a "true copy" which preserves content or meaning |
| Accurate | Free from errors |
| | No editing performed without documented amendments |
| | Conforming to truth or standard |
| Consistent | All data, and relevant metadata, including any repeat or re-analysis performed |
| Enduring | Application of good documentation practices throughout any process |
| | The application of date and time stamps in the expected sequence |
| Available | Available and accessible for review, audit, or inspection throughout the retention period |
| | |
| *Traceable* | *Changes to the data, to the context/metadata are traceable throughout the life cycle of the data* |
| | *Changes documented as part of the metadata* |

Data integrity expectations and holistic approaches are not new in the global regulation scene. EU GMP 1.4, US GMP 21 CFR 211.160 (b), and US GMP 21 CFR 211.194 (a), all have insinuations to it. The aim of the ISPE guideline on computerized systems is to provide practical "bridge" between achieving compliant GxP computerized system and good data governance for regulated business processes. Data integrity by design is quality critical initiative and it is created to ensure public health, patient safety, and product quality in GxP regulated environment. To achieve this governance framework is combined with Quality Risk

Management (QRM) and Knowledge Management (KM). Data integrity, patient safety, and product quality are achieved by collectively applying QRM, which in turn acts as an input to KM. This drives continual improvement to business and organizational processes. (*ISPE Data Integrity by Design, 2020, pp. 9 & 15.*)

**The Reference Architectural Model for Industrie 4.0** (RAMI 4.0) is based on the IEC 62890, IEC 62264 and IEC 61512 standards which refer to life-cycle management, enterprise-control system integration and batch control (*RAMI 4.0 | Isa.Org, n.d.*). The three-dimensional map generated from the three standards demonstrate the most important aspects that the 4IR has and its developed to ensure common perspective and understanding amongst all participants (*Reference Architectural Model Industrie 4.0 | Plattform Industrie 4.0, 2018*).

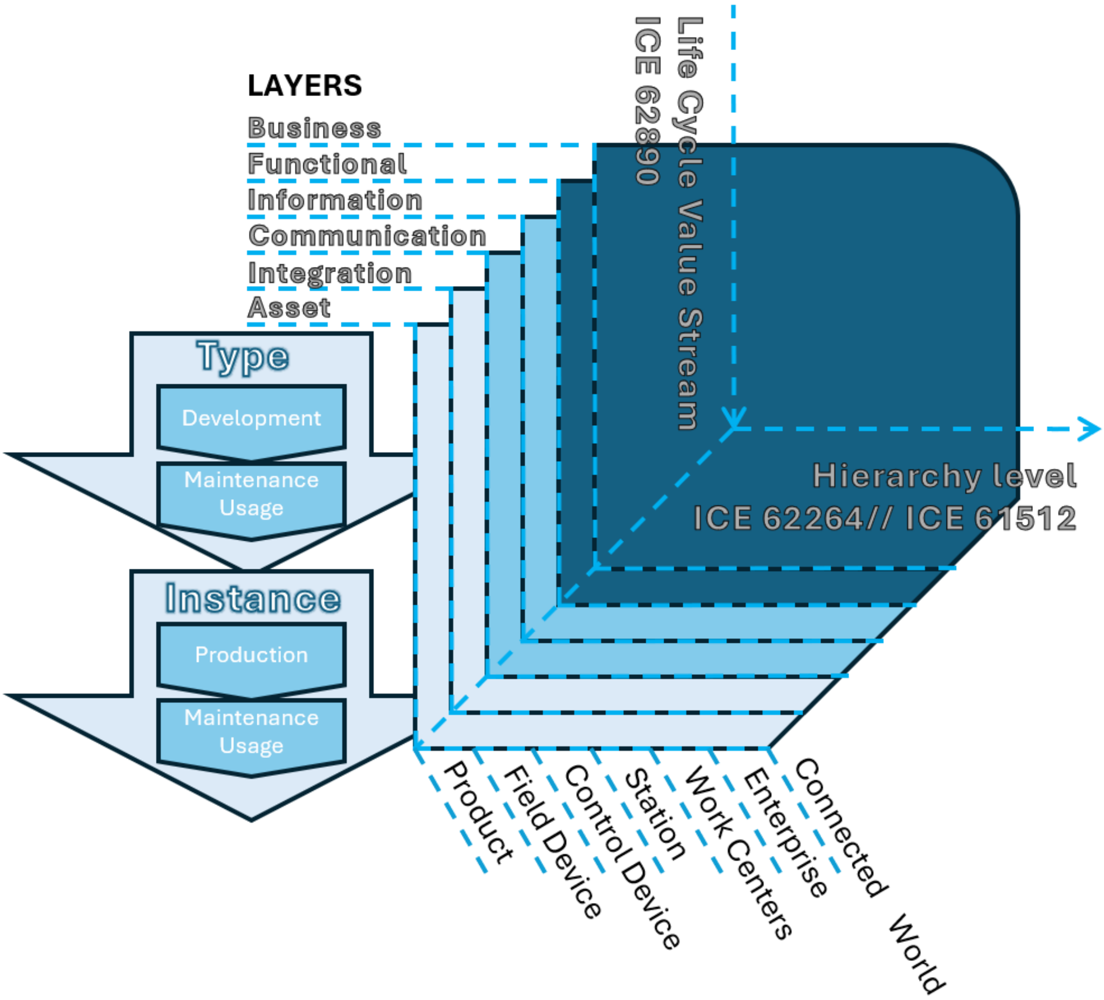



Figure 3. RAMI 4.0 reference architecture (adapted from *Reference Architectural Model Industrie 4.0 | Plattform Industrie 4.0, 2018*).

Rami 4.0 can be used by companies to develop future business models and products by using a common framework which allows complex processes to be broken down into easy-to-grasp blocks. The service-oriented architecture of RAMI 4.0 presented in Figure 3 describes all of the crucial aspects of 4IR in its three-dimensional coordinate system. This allows breakdown of complex interrelations into simpler and smaller clusters. (*RAMI 4.0 | Isa.Org, n.d.*.)

The three axes are axis 1 - hierarchy, axis 2 - product life cycle and axis 3 - architecture. In Table 2 the hierarchy levels of axis 1 come from IEC 62264 standard that describes enterprise IT and control systems. Axis 2 is based on IEC 62890 standard where the life cycle of products and facilities are represented. Axis 3 has 6 layers of machine layers that are decomposed into its properties. This type of representation originates from information and communication technology, and it is commonly used to break down complex systems into layers. (*RAMI 4.0 | Isa.Org, n.d.*.)

Table 2. Three axes of RAMI 4.0 (*Reference Architectural Model Industrie 4.0* | *Plattform Industrie 4.0, 2018*).

<table>
<tr><th colspan="2">Axis 1 - Hierarchy</th><th colspan="2">Axis 3 - Architecture</th></tr>
<tr><td colspan="2">Flexible systems and machines</td><td>Business</td><td>Organization and business process</td></tr>
<tr><td colspan="2">Functions are distributed throughout the network</td><td>Functional</td><td>Functions of assets</td></tr>
<tr><td colspan="2">Participants interact across hierarchy levels</td><td>Information</td><td>Necessary data</td></tr>
<tr><td colspan="2">Communication among all participants</td><td>Communication</td><td>Access to information</td></tr>
<tr><td colspan="2">Product is part of the network</td><td>Integration</td><td>Transition from real to digital world</td></tr>
<tr><td colspan="2"></td><td>Asset</td><td>Physical thing in the real world</td></tr>
<tr><th colspan="4">Axis 2 - Product life cycle</th></tr>
<tr><td>Development<br>-Construction plan:</td><td>Maintenance usage<br>-Construction plan:</td><td>Production<br>-Production:</td><td>Maintenance usage<br>-Facility management:</td></tr>
<tr><td>Development</td><td>Software updates</td><td>Product</td><td>Usage</td></tr>
<tr><td>Construction</td><td>Instruction manual</td><td>Data</td><td>Service</td></tr>
<tr><td>Computer simulation</td><td>Maintenance cycles</td><td>Serial number</td><td>Maintenance</td></tr>
<tr><td>Prototype</td><td>…</td><td>…</td><td>Recycling</td></tr>
<tr><td>…</td><td></td><td></td><td>Scrapping</td></tr>
<tr><td></td><td></td><td></td><td>…</td></tr>
</table>

## 2.2 Cloud Computing

Cloud computing paradigm is a model (not a specific technology) for delivering computing services over the internet (public or private), where data, applications, and resources are hosted and managed on remote servers (data centres) rather than on local systems e.g., On-Premises data centre. Cloud computing infrastructure covers hardware and software, and it provides on-demand access to shared pool of computational resources like storage, processing power, and software via the internet, enabling scalability, flexibility, and cost-efficiency. Cloud Computing allows organizations to scale their need, access services globally, and benefit from centralized data management and analytics. It is widely used for applications that require extensive computing power, large-scale data storage, and remote access, such as data analysis, web hosting, and enterprise software solutions. (*Jackson & Goessling, 2018; Mei, 2022; What Is Cloud Computing Infrastructure | VMWare, 2024*)

### 2.2.1 Definition and key concepts

In terms of definition, cloud computing can be expressed as services, hosted by service provider, that are accessed on-demand, and the shared pool of computing resources are fully configurable for the needs of the user (*Mell & Grance, 2011, 2*). National Institute of Standards and Technology also list essential characteristics, service models and deployment models that their cloud model consists of. These three groups and their contents are presented in the Table 3 below.

Table 3. NIST cloud model (*Mell & Grance, 2011, 2-3*).

| ESSENTIAL CHARACTERISTIC | DEPLOYMENT MODELS | SERVICE MODELS |
|---|---|---|
| On-demand self-service | Private cloud | Software as a Service (SaaS) |
| Broad network access | Community cloud | Platform as a Service (PaaS) |
| Resource pooling | Public cloud | Infrastructure as a Service (IaaS) |
| Rapid elasticity | Hybrid cloud | |
| Measured service | | |

To further examine these characteristics and models, Mell and Grance (*2011*) of the NIST, open the definitions of each. For the better understanding of this thesis and the technological aspect of it, few of the most significant definitions are introduced next.

**Resource pooling:** Cloud systems providers computing resources are pooled to be used by multiple different customers at the same time. The resources can be different virtually and physically. Usually, the customer does not  have sense of the physical location, and the virtual characteristics dynamically change. Examples of resources include but are not restricted to storage, processing, memory, and network bandwidth.

**Rapid elasticity:** Capabilities can be elastically provisioned and released to the customer. To them, the capabilities available often appear to be unlimited and are available at any time.

**Measured service:** Cloud systems dynamically control and optimize resource usage. This usage can be for example monitored, controlled, and reported. Both the provider and the customer benefit from the transparency this gives to them.

**Software as a Service (SaaS):** Customer has the capability to utilize the service provider's applications that are run on the cloud infrastructure. The customer can access these applications via e.g., web browser or program interface. The service provider manages the cloud infrastructure as a whole and the customer does not need to worry about network, servers, storage or anything infrastructure related. Not even individual application capabilities, with exception of finite user specific configuration settings in the application.

**Platform as a Service (PaaS):** Customer has the capability to deploy onto the cloud infrastructure customer-created and/or acquired applications that are built using programming languages, libraries, services, and tools supported by the service provider. The service provider manages the cloud infrastructure as a whole and the customer does not need to worry about network, servers, storage, but the customer has control over the deployed applications and in some cases, the configuration settings for the application-hosting environment.

**Infrastructure as a Service (IaaS):** Customer has the capability to provision processing, storage, networks, and other basic computing resources where the customer is able to deploy and run wanton software. These can include e.g. operating systems and other applications. The service provider manages the cloud infrastructure as a whole and the customer does not need to worry about network, servers, storage, and deployed applications; and the customer possibly has limited controls over other networking components (e.g., host firewalls).

**Private cloud:** The cloud infrastructure is dedicated to exclusively for use by a single organization that can consist of one or more customers (e.g., production facility). It can be owned, managed, and operated by the organization itself, a third-party IT consult, or some combination of these, and it can be located on or off premises.

**Public cloud:** The cloud infrastructure is intended for open use by the general public. It can be owned, managed, and operated by an individual company, academic entity, or government organization. It fully exists on the premises owned by the cloud provider.

**Hybrid cloud:** The cloud infrastructure is combination of two or more distinct infrastructures that are combined by technology that is standardized or proprietary, and which enables portability of data and applications.

These NIST definitions apply to cloud computing as a whole and they are not uniquely written to pharma environment. The regulation and legislation regarding pharma industry sets the bounds to implementing cloud services in the manufacturing processes and enabling functions. For example, ISPE GAMP® 5 Guide, 2$^{nd}$ edition considers the validation of SaaS solutions that are previously not covered by the older edition. It also states that if any cloud resources are used, the qualification of the infrastructure needs to be considered. (*ISPE GAMP® 5, 2022.*)

GAMP® 5 (*2022*) visualizes the differences between traditional IT, IaaS, PaaS and SaaS implementations regarding the responsibilities of supplier managed and customer managed parts of the implementation. This visualization can be seen in Table 4 below.

Table 4. GAMP® 5 Responsibilities of supplier and customer (adapted from *ISPE GAMP® 5, 2022, 169*).

| | Traditional IT | IaaS | PaaS | SaaS |
|---|---|---|---|---|
| Applications | X | X | X | O |
| Data | X | X | X | O |
| Runtime (RT) | X | X | O | O |
| Middleware | X | X | O | O |
| Operating System (OS) | X | X | O | O |
| Visualization | X | O | O | O |
| Servers | X | O | O | O |
| Storage | X | O | O | O |
| Networking | X | O | O | O |

| | |
|---|---|
| X | = Customer manages |
| O | = Supplier manages |
| | = Infrastructure |

From this table, it can be observed that the infrastructure is mostly managed by the supplier of the cloud solution, except in IaaS, where the OS is managed by the customer. The landscape of “- as a Service” is grown largely and big tech companies providing these services, have marketed the acronym XaaS – anything as a service to include solutions others than the three presented before, such as Containers as a Service (CaaS), Function as a Service (FaaS) and Artificial Intelligence as a Service (AIaaS) (*What Is AIaaS | Microsoft Azure*, n.d.; *What Is XaaS | IBM*, n.d.).

Regarding whether to use public or private cloud, GAMP® 5 has nothing to help with the decision. It only mentions that when configuration and design is under work, the architecture of the hardware design of the computer system should cover public/private cloud and its storage capacities (*ISPE GAMP® 5, 2022, 201*).

About the Essential Characteristics of the cloud model described in the NIST definition by Mell and Grance (*2011*), GAMP® 5, Annex 11 or such other does not directly cover the characteristics of Cloud solutions but these are parameters that affect the decision-making process regarding implementation of cloud and the Service Models used in pharma environment. To further investigate the impact on the decision-making process, the characteristics of the cloud are covered in Chapter 2.2 where the significance and relevance of each is listed and then evaluated in the context of pharmaceutical environment.

The key concepts and the significance of cloud solutions come from the flexible subscription type services or in “pay-as-you-go" basis. This enables the end users to access services such as SaaS and IaaS with low budget where they only pay for the e.g., the applications and storage they are using. Other significant factors with Cloud services are scalability and flexibility in resource management, high availability and reliability of the infrastructure and services, automated updates and maintenance, cost efficiency that was previously described with the pay-as-you-go basis, and global access and easy collaboration within the organization. (*Chaudhary et al., 2017.*)

Real life example of benefiting from using cloud services in pharmaceutical industry is implementation of a cloud-based method for protected clinical trial data sharing. In this article, Luthria and Wang have used open-source data analysis tool that is utilized for cloud computing and users can execute flexible queries from their personal computers that runs on the cloud. They demonstrate real data summary statistics and synthetic data to analyse clinical trial data. Data privacy aspects are mitigated by restricting types of queries users can run. (*Luthria & Wang, 2020.*)

Another example of getting benefits from using cloud computing, is using machine learning and artificial intelligence in drug discovery. As the amount of data processed (and used in training data) increases, and the AI/ML models get more complicated regarding complexity of computational structures, the need for larger and more powerful resources e.g., memory, storage, Central Processing Unit (CPU) and Graphics Processing Unit (GPU) increases (*Mohammed et al., 2023; Reddy, 2019; System Requirements for Deep Learning | GeegsforGeeks, 2024; Tu et al., 2024*). Spjuth, Frid and Hellander describe how ML used in cloud computing can help the ML life cycle in drug discovery. They go through how cloud computing can ease the challenges related to ML modelling and they discuss on how private and sensitive data and cloud environments are working together. They also state that AI has become important tool in today's drug discovery and that cloud computing presents many tools that are helpful regarding the use of AI and ML in drug discovery. Despite cloud solutions having many advantages, there is always cost and data privacy effects when using public cloud services. (*Spjuth et al., 2021, 1077.*)

### 2.2.2 Evolution and growth

Early roots of Cloud Computing can be traced back to as early as the mid-1940's and as late as 1990's, depending on the sources (*Maayan, 2021; Who Coined "Cloud Computing"?, 2011*). The exact year of the phrase "Cloud Computing" is irrelevant to this research and for this reason the history, evolution and future of the paradigm is briefly and broadly gone through. The division of events is done to five stages, foundation, emergence of modern cloud, cloud adaption explosion, present day (2020 and forward) and market trends and forecast.

To list significant milestones in the path to today's cloud computing there are few notable years and projects that frequently appear in the literature.

Foundation:

- Time-Sharing mainframe by John McCarthy **1960's** (*Reminiscences on the Theory of Time-Sharing, n.d.*)

- Advanced Research Projects Agency networks (ARPANET) by J.C.R Licklider **1960**'s (*Packard, 2023*)

- Virtualization software as an alternative to Time-Sharing **1970**'s (*IBM Virtual Machine Facility/370, 1973*)

Emergence of modern Cloud:

- World Wide Web (WWW) by Tim Berners-Lee **1990's** (*Sarafe, 2023*)

- Phrase "Cloud Computing" in academical publishment **1990's** (*Foster & Kesselman, 1999*)

- Rise of SaaS and PaaS **2000's** (*Matana, 2024*)

- AWS Simple Storage Service (S3) and Amazon Elastic Compute Cloud (EC2) **2006** (*Choudhari, S. et al., 2021; Choudhary, A. et al., 2021*)

- Google App engine launch **2008**, PaaS solution (*Reflecting on Our Ten Year App Engine Journey, 2018*)

Cloud adaption explosion:

- Growth of high-speed internet, virtualization technologies, and scalable infrastructure led to widespread cloud adaption **2010's** (*How Tech Evolved in the 2010's | Global X, 2020; The History of Virtualization and Its Mark on Data Center Management | TechTarget, 2019; "Top Cloud Computing News Of The Decade | VMWare," 2021*)

- First hybrid cloud **2011** (*Deb & Choudhury, 2021*)

- Multi-cloud and hybrid cloud advancements **2010's** (*Deb & Choudhury, 2021; Ferrer et al., 2016*)

- Containers and Kubernetes **2014** (*The Evolution of Containers | TechTarget, 2023*)

- Huge market size growth **2010's** through **2020's** (see Appendix 1 Trend in 2010's and 2020's for cloud market)

Present day:

- Ubiquity of cloud services (*Vahdat Nejad et al., 2018*)

- Emerging technologies driving cloud evolution (*Eight Emerging Trends Shaping The Future Of Cloud Computing | Forbes, 2024*)
- Regulatory and compliance enhancements (*The Power of Choice, 2024; Najana & Ranjan, 2024*)
- Cloud in pharmaceutical industry (*Cloud Computing in Pharma | Viseven, 2024*)

Market trends and forecast:

- $1T market by 2027 (*The Future of Cloud 2028 | Gartner, n.d.*)
- Quantum computing (*Golec et al., 2024*)
- Growth of AI and ML applications use, and development (*Top 10 Cloud Computing Trends | GeeksforGeeks, 2024*)
- IoT (*The next Generation Internet of Things | European Commission, n.d.*)
- Blockchain (*The Future of Cloud Computing 2025-2030 | CloudDefense.AI, 2024*)
- Edge and cloud merging (D'Agostino et al., 2019; *Merging Cloud with Edge Computing | LinkedIn, 2024*)
- Security development (*The Future of Cloud Cybersecurity | CSA, 2024*)
- "Finance" + "DevOps" (FinOps) (*Using FinOps to Optimize AI and Maximize ROI | DataScienceCentral, 2025*)

Cloud computing is not a new concept, the paradigm roughly 60 years old but commercially cloud services have been available since 1999 when Salesforce offered first applications over the internet (*The History of Cloud Computing | Cyber Magazine, 2021*). In these rough 60 years the cloud computing has evolved to its present form and popularity. Many big tech companies are heavily involved in the development of different cloud solutions, and the research is massive business. In the history between 2010's and 2020's, the rapid market growth has been substantial (see Appendix 1 on page 94) and the estimation of the market size by 2027 is braking 1 trillion USD (*The Future of Cloud 2028 | Gartner, n.d.*). And as the trendlines show in Figure 20, Figure 21 and Figure 22,cloud computing has been in significant popularity amongst global corporations as the spending has been increasingly growing for decades.

### 2.2.3 Benefits and limitations

Cloud computing brings many benefits to the table for businesses to take advantage on, and for them to utilize it and create more efficient ways of doing business. Cloud is not perfect solution and therefore also the limitations that it has are studied in this thesis. For businesses to successfully utilize cloud computing it requires in-depth financial and technical analysis before the actual implementation of architecture and applications can start (*Rosati & Lynn, 2020, p. 20*). In 2008 IBM conducted private study that was aimed for business executives representing multiple different enterprises to identify the main business benefits of cloud computing. Four benefits from using cloud emerged from the survey results; accelerated business innovation, accelerated technology deployment, reduced IT management costs and reduced capital costs. (*Furht et al., 2010, p. 66.*) These benefits are still valid to this day.

The benefits that cloud has to offer typically relate to cost savings in the form of reduced total cost of ownership, especially cutting the initial investment down, when no HW or SW costs come from deployment when using public cloud (*Rountree & Castrillo, 2013, p. 37*). Cloud computing advantages can be summarized into four main points; on-demand computing recourses that are basically unlimited, IT architecture and available HW & SW resources that are instantly available, pricing based on usage and processing environment that is troublesome to generate using single machine (*Huawei Technologies Co., 2022, p. 22*). Other benefits of cloud include e.g., resource utilization and management, service availability and reliability, cost effectiveness and better security (*Furht et al., 2010, p. 256*).

Table 5 shows cloud benefits according to Chopra (*2017*) and better description and also the impact of each is mentioned that are found in literature (*Antonopoulos & Gillam, 2017; Chopra, 2017; Furht et al., 2010; Huawei Technologies Co., 2022; Lynn et al., 2020*).

Table 5. Cloud benefits and their descriptions.

| **Benefit** | **Description** |
|---|---|
| Resource management | On-demand and dynamic provisioning of resources and applications, such as CPU, memory, storage and software that run processes between client and server.<br>Including monitoring of resource usage and managing capacity of resources.<br>Contributes largely on the optimization of performance and cost. |

| Benefit | Description |
| --- | --- |
| Dynamic scaling | Ability to dynamically scale in and out the application depending on the demand of resources and spikes that occur in traffic.<br>Has cost impact by reducing resources when load is smaller and not requiring overprovisioning. |
| High availability and durability | Service Level Agreement (SLAs) define the level of availability and durability of data, and with big service providers, the number of nines (higher the number the better given attribute is) is big resulting in reliable, robust and accessible storage.<br>Enabled via e.g., replication techniques and redundancy.<br>Gives business value with reliability and risk mitigation. |
| Highly available services | High availability of services such as critical systems and applications is always up and accessible for users. Enabled with redundant servers and automatic failover protection.<br>This protects businesses revenue stream, guarantees customer satisfaction and improves productivity. |
| Periodic workloads | Applications that are only made available periodically and after its function is fulfilled, they are removed and costs generated of the use of the application stop. For example, loads requiring large computational power that are used few times a week are not costing resources or money when they are not utilized.<br>Has cost impact. |
| Unpredictable growth | No need to overprovisioning or danger of underutilization when cloud environment is dynamically scaling. Business growth does not need to be predictable for the systems to meet requirements.<br>Guarantees customer satisfaction and enables business growth w/o needing to overprovision infrastructure.<br>Brings cost savings by deleting the need to build infrastructure in advance and gives flexibility for the business. |
| Workload spikes | With autoscaling (previously explained dynamic scaling) cloud computing takes pressure of the situations where the servers experience abnormally high demand. No fear of losing customers or revenue due to workload spikes disrupting applications/environment.<br>Cost benefits when there is no need for prepare for unexpected workload spikes by oversizing infrastructure. |
| Infrastructure offloading | Migration of on-premises networks, servers and applications to the cloud.<br>Helps understand the cost structures better with service providers providing transparent pricing models. No need for manage own datacentres or servers that generate expenses in terms of workers needed, energy consumed, hardware utilized, and software and licences needed. |

Additionally cloud computing gives advanced security benefits for the system with deep and breadth security features that are monitored and developed around the clock preventing catastrophic events when data is compromised (*Antonopoulos & Gillam, 2017; Lynn et al., 2020*). According to Kappelman et al. (*2017*) cybersecurity is amongst the largest IT spending category in 2017 for organizations and with cloud computing – businesses are able to cut costs and make sure that the security measures are top tier (*Advantages Of Cloud Computing* |

*Google Cloud, n.d.; Huawei Technologies Co., 2022; Tallon et al., 2020, p. 63*). Gorton, Greenfield, Szalay and Williams discuss research issues regarding data/compute-intensive problems in 21st century and these fundamental challenges are developing new algorithms that scale to process and search exponentially growing data, substantially reducing associated data analysis cycles to uphold practical, topical applications as well as processing and managing massive amounts of growing data volumes (*Gorton et al., 2008*). These fundamental challenges can be addressed with cloud computing and its scalable resources (*Antonopoulos & Gillam, 2017*).

The limitations that are associated with cloud computing relate to e.g., relying on internet access, less control over the cloud infrastructure and integration related challenges. Cloud computing also poses businesses to additional privacy, latency, bandwidth and scale challenges. According to Shi et al. (*2016*) the growing amount of data created at the edge devices is causing bottleneck for cloud solutions regarding transportation speed of data.

The implementation and transformation from traditional on-premises datacentres to cloud environment needs to be properly done in order to avoid negative impacts to the business. (*Advantages Of Cloud Computing | Google Cloud, n.d.; Furht et al., 2010, pp. 74–75.*) Challenges that cloud computing faces are listed in Table 6 and small description of each is added to explain each.

Table 6. Cloud challenges and their descriptions according to (*Furht et al., 2010, pp. 17–18*).

| **Challenges** | **Description** |
|---|---|
| Bandwidth cost | Implementing cloud solutions, companies can save money on HW and SW, but they are being charged over the use of network bandwidth. Data intensive applications result in larger costs. |
| Control | Cloud computing platforms for individual companies and the providers have full control over their systems. |
| Performance | Cloud computing lacks adequate performance when intensive transaction oriented and alike data-intensive applications need to be run. Physical location (far away from source) causes high latency and delay. |
| Reliability | Outages are high risk for companies that rely on round-the-clock reliability. |
| Security and privacy | Critical IT resources being outside companies own firewalls rais security and privacy worries. Solution is to not use public and shared environments but private cloud environments to meet SLAs. |

Considering challenges that need to be addressed in the pharmaceutical working environment, security and privacy is big concern. Studies and development of cloud-based methods for protected data sharing have been done see e.g., (*Luthria & Wang, 2020*), where they have developed protected method for data sharing which is fundamental requirement in pharma. Considering the regulatory requirements and law, for example GMP, General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), that require pharma industry to handle Personally Identifiable Information (PII) and data safely and securely (*EudraLex - Volume 4 - European Commission, 2025; Health Insurance Portability and Accountability Act of 1996*, n.d.; *Regulation - 2016/679 - EN - GDPR - EUR-Lex, 2016*). It is in high priority for pharmaceutical companies that the data they produce, and use is protected, and they comply with the regulations.

### 2.2.4 Architecture

Different cloud providers have different architectures in their products see e.g., (*Furht et al., 2010, pp. 343–345*). In more universal level, the cloud computing architecture can be viewed as seen in Figure 4, where the architecture is divided into two parts, frontend and backend, each of these is connected to the internet. The frontend is the client's side e.g., application interface, and the backend is the service provider's end. Client-side infrastructure element provides the graphical interface for the client to cloud. Internet element between frontend and backend is a medium which enables the communication between each side. Application element is the software that client wants to get access to. Service element is for example IaaS, PaaS or SaaS, there are also other services available nowadays. Runtime cloud element enables execution and RT environment to virtual machines. Storage element is used to store and manage data, and it is one of the most important parts of cloud architecture. Infrastructure element in the backend side offers services on application, host and network levels. Management element acts as the backbone of the architecture, managing all of the other elements at the backend side. Security element deals with security at the backend. (*Shapla et al., 2020, p. 838.*)

For the whole implementation of cloud computing solution to be most valuable and efficient for the business, the decision of the technology to be used and the service provider needs to be done, also the functional requirements need to be answered. These are core business drivers (e.g., cost savings and scalability), performance requirements, and data sovereignty and compliance requirements. Deployment model and service model needs to be decided, whether it is going to be private, public or hybrid cloud and whether it is going to be IaaS, PaaS or SaaS (see Chapter 2.2.1 for more thorough explanation of these). (*Gaca, 2023.*)

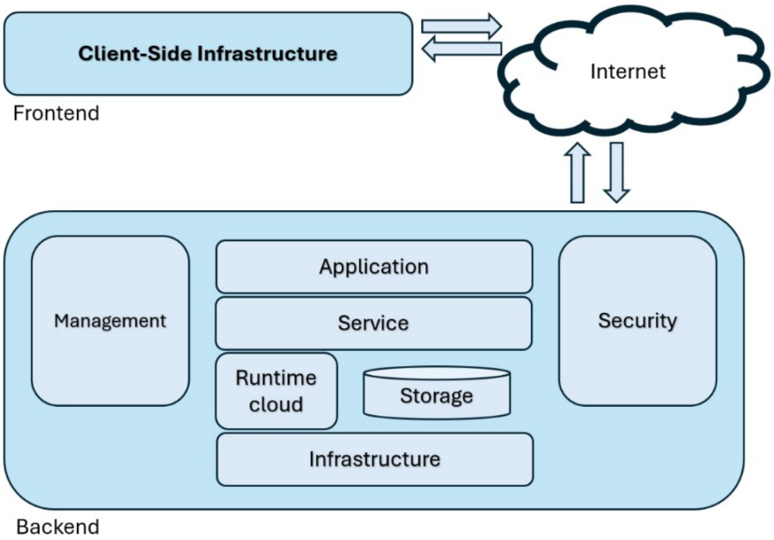



Figure 4. Cloud computing architecture (Adapted from *Shapla et al., 2020*).

In Figure 5 Cloud computing reference architecture by NIST is presented showing major actors, their functions and activities in a clear manner. This architecture is generic and high-level presentation, and it is designed to help reader understand the characteristics, requirements, standards and uses of cloud computing. In this architecture, NIST defines five major actors cloud auditor, broker, carrier, consumer and provider. Each of these is either a person or an organization that is involved in the transactions or process and/or performs activities in cloud. (*Liu et al., 2011, p. 3.*)

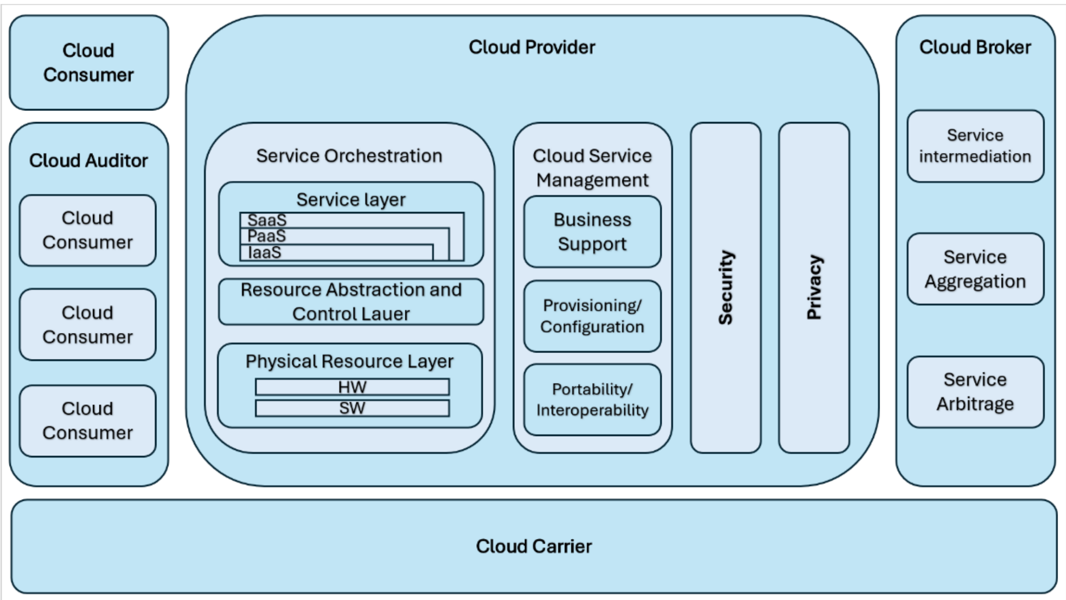



Figure 5. Cloud computing reference architecture by NIST (Adapted from *Liu et al., 2011, p. 3*).

To define the five actors presented in Figure 5 Liu et al. (*2011*) have formed a clear and brief table to easily recognize each actor's main characteristics. According to their definition,

- **Cloud auditor** is able to independently perform assessment of cloud services, and operations, security and performance of the information system.
- **Cloud broker** manages delivery, performance and use of cloud services. It is also responsible of the negotiation of relationships between cloud consumer and cloud provider.
- **Cloud carrier** provides connectivity and transport of cloud services to cloud consumers from cloud providers.
- **Cloud consumer** is a person or organization, and they maintain business relationship with cloud providers and use their services.
- **Cloud provider** is either person, entity or organization that produces and delivers services to the customer.

NIST reference architecture is vendor-neutral and designed to the need of described overall framework. The architecture tries not to stifle innovation by defining prescribed technical solution. It lays out the central elements of cloud computing and it crates level playing field for the cloud providers to develop their products according to this Actor/Role based model. (*Liu et al., 2011, p. vi.*)

## 2.3 Edge Computing

Edge computing is a distributed, decentralized computing paradigm that extends cloud computing services to the edge of the network by processing data closer to the source where it is generated (e.g., IoT devices, sensors and local servers) rather than relying on centralized cloud data centres that are possibly on the other side of the world. This proximity reduces latency, improves real-time responsiveness, enhances data privacy, and minimizes bandwidth usage, making it ideal for applications requiring immediate analysis or compliance with strict data residency laws. (*Taheri & Deng, 2020; Yeung, 2019.*)

### 2.3.1 Definition and key concepts

In terms of definition, edge computing is technology that enables computation at the edge of the network. Downstream data comes from cloud and upstream data comes from the IoT devices and services. Reason for edge computing is to get data processing close to the data source. (*Cao et al., 2018, p. 2; Shi et al., 2016, p. 638.*) In other words, edge can be described as optimising model of cloud system (*Taheri & Deng, 2020, p. 8; S. Singh, 2017*). Major drivers that propel the shift to edge comes from the following requirements. A push from cloud-side towards solution that fixes the draw backs of the cloud, e.g., latency and bandwidth issues. A pull from IoT-side that most electric devices will be connected to it in the near future, which means that the devices produce massive amounts of data, and the devices also consume data making the cloud solution inadequate. Shift to data producing from data consuming, changing common cloud solution where data is consumed at the edge device, data in the cloud is sent to the edge device (downstream), to the edge device producing the data and sending it towards the cloud and/or edge layer (upstream). Decentralizing cloud solutions for geographically distributed applications and enable low-latency computing by bringing the computation closer to the source of the data. Overcoming the resource limitations of edge devices and limit the amount of data sent to the cloud to be processed by using edge nodes to filter, analyse and compute the data before sending it to the cloud. Sustaining reasonable levels of energy consumption limiting the excess use of resources in the cloud which energy demand is becoming unsustainable. Coping with network traffic and explosively growing amounts of data which cloud architectures are not designed to handle especially regarding variety, volume and speed of which the data is being generated. Lastly a need for manipulation of network resources and the distribution of computation between these resources via smart computation techniques such as edge nodes being able to facilitate computation close to data source and being able to enhance edge devices capabilities remotely. (*Al-Dulaimy et al., 2020, pp. 4–7.*)

Edge components are visualized in the Chapter 2.3.4 and Chapter 2.6 where the architecture is presented more thoroughly. Main components are the edge devises (IoT) which are the sensors and equipment that create the data (data source), edge servers that are located in the proximity of the end user and which take care of the heavy computing, edge routers handle data transmissions across networks, edge switches connect the edge devices together within the local network, and edge nodes and gateways filter, process and sometimes locally store data before it is sent upstream to the cloud (*K. Cao et al., 2020*).

Key concepts and the significance of edge computing is seen as quite straight forward in the literature – to provide smart services at the proximity of the edge of the network, migrate cloud capabilities to the edge, meet agile and flexible needs with responsive connectivity and integration between end-to-end systems and their components, and addressing the requirements of latency, bandwidth, data optimization, security and privacy. Other significant factors of the edge are scalability, energy efficiency, decentralization and context awareness. (*K. Cao et al., 2020; Mannanuddin et al., 2020; Shi et al., 2019; Taheri & Deng, 2020.*)

Real life example of benefiting from using edge computing in healthcare industry is implementation of computing at the edge of the network in medical data collection and processing. This work on edge-based smart healthcare system is done by Nan et al. (*2020*) in (*Taheri & Deng, 2020, pp. 337–356*) and it proposes three-layered (see Figure 6) implementation of patient-oriented healthcare system which is enabled with edge computing. They present successful case studies of daily monitoring and early warning of diseases such as diabetes, cardiovascular diseases and neurological diseases where edge computing has been implemented and utilized, and they also present valid open issues associated with the implementation of this type of edge-based systems.

Another example of the benefits of edge-based solution is presented by Shi et al. (*2016*), where they present multiple case studies where they argue that edge computing could shine. For this example, the most interesting case study is the cloud offloading that Shi et al. have brought up first. As they say and as it is also presented in Chapter 2.2 in this thesis, in cloud computing paradigm, the computation happens primarily in the cloud and the issues this might present are e.g., issues with latency which can drastically lower the experience that the user has. In this case study example, Sie et al., present the reader few non-pharma related cases where the user could benefit from the computation being done closer to the end-user, by offloading the processing from cloud to edge nodes and therefore reducing the latency. They also list content filtering/aggregation as one beneficiary application of the cloud offloading, and this is also one that is quite interesting in relation to this thesis and the regulated working environment that is being studied here.

### 2.3.2 Evolution and growth

Need for edge computing has risen from the growing number of IoT and smart devices connected to network, and the exponential growth of data produced by these devises and the need of processing this data quickly and effectively. Before the term “edge computing” was introduced first time in 2002, the era of when cloud computing was flourishing. (*Al-Dulaimy et al., 2020, pp. 3–4; Mannanuddin et al., 2020, pp. 1–2; Margulius, 2002.*)

In the literature, edge computing and its history is presented in different ways depending on what each writer wants to highlight but the big milestones are always the same. For this thesis, the division is done to five segments, foundation, emergence of modern edge, edge adoption explosion, present day (2022 and forward) and market trends and forecast. Mannanuddin et al. (*2020*) presents edge computing as the "fourth computing wave", third being cloud computing, second being the technology of web and the first being monolithic systems. Other way of presenting the evolution is to divide the history into three stages, technology preparation period, rapid growth period, and steady development period (*Shi et al., 2019, pp. 1476–1477*).

To list significant milestones in the path to today's edge computing there are few notable years and projects that frequently appear in the literature.

Foundation:

- Content delivery networks (CDNs) **1990's** (*George & George, 2021, p. 3*)
- Emergence of IoT **1999-2010's** (*Braun, 2019*)
- Introduction of fog computing **2010's** (*Kaswan et al., 2024, pp. 1–2*)
- Cloud computing emergence **2010's** (*Foote, 2021*)
- Edge computing concept **2012** (*The Rise of Edge Computing | ATP.Org, 2024*)

Emergence of modern Edge:

- European Telecommunications Standards Institute (ETSI) Mobile Edge Computing (MEC) standard **2015** (*Patel et al., 2015*)
- OpenFog consortium **2015** (*OpenFog | OPCfoundation.Org, n.d.*)
- Edge computing consortium **2016** (*Edge Computing Consortium Is Established to Deepen Digital Transformation - Huawei Press Center, 2016*)

Edge adoption explosion:

- Edge AI emergence (*R. Singh & Gill, 2023*)
- Cloud-Edge integration (*How Integrating Edge with Cloud, Data and AI Unlocks Value, 2023*)

Present day: (*Investing in Cloud, Edge and the Internet of Things | Shaping Europe's Digital Future, n.d.; R. Singh & Gill, 2023*)

- Industrial and government adoption
- Edge-Cloud continuum
- Maturation of Edge AI

Market trends and forecast: (*Investing in Cloud, Edge and the Internet of Things | Shaping Europe's Digital Future, n.d.; R. Singh & Gill, 2023*)

- Secure and sovereign edge
- Autonomous edge orchestration
- Sustainability and energy efficiency
- Integration with Industry 4.0 and digital twins

### 2.3.3 Benefits and limitations

Edge computing has many benefits, and it has been developed to make up for the shortcomings of cloud computing, such as latency and bandwidth. Although edge offers businesses major benefits and acts as driver towards innovation e.g., smart manufacturing, autonomous vehicles and real time decision-making it still has its own limitations such as processing power and storage capacity. (*Shi et al., 2019.*) These benefits and limitations are studied in this chapter.

Table 7 and Table 8 list key benefits and limitations associated with edge computing and their description according to multiple publications (*Bhattacharya, 2022; K. Cao et al., 2020; Hoareau & Satoh, 2009; Maciel et al., 2022; Mannanuddin et al., 2020, pp. 4–5; Shi et al., 2016, p. 639; S. Singh, 2017, p. 52; Taheri & Deng, 2020, p. 10*). With edge computing, the processing can be brought closer to the data source to reduce e.g., latency, lower bandwidth consumption, and real-time processing and analysis of data. With these benefits and others seen in Table 7, edge computing offers great solution for businesses.

Despite all of the benefits edge computing brings, it still has limitations that need to be addressed, and they affect the decision-making process. Processing power of edge devices is limited and that affects the type and amount of processing it is possible to be done in these. Devises that are commonly limited with processing power are gateways, microcontroller, etc. Network complexity arises as challenge when distributed system is large, and it has many layers and devices connected to it have diverse capabilities which are not allocated properly. Table 8 also show that the system constrains might result in losing some benefits when for example Availability and reliability is prioritized, then benefits in energy efficiency are limited or lost completely. (*K. Cao et al., 2020; Pérez et al., 2022; Shi et al., 2016.*)

Table 7. Edge benefits and their descriptions

| **Benefit** | **Description** |
|---|---|
| Close to real-time analysis of data | Data analysis is done close to real-time at the proximity of the data source and not on distant cloud or data centre which is lot slower. |
| Context-awareness | Is a system that can adjust itself based on the conditions of the system. This helps edge nodes make smarter and faster decisions based on e.g., devise state and network conditions. |
| Compliance | Edge devices meet regulatory requirements regarding e.g., audit trail, data integrity, and data portability capabilities. |
| Energy efficiency improvement | Improved energy efficiency and scalability for IoT devices. As a result of this also the physical size of the devices can be reduced. |
| Latency reduction | Reduction on response time and time to action, also network resources are conserved. |
| Low bandwidth | Local processing of data reduces the burden of bandwidth that is higher if lots of data needs to be transmitted upstream to cloud/data centres. |
| Network load reduction | Reduction of data flows in the network due to smaller amount of data being transmitted over to cloud/data centres resulting in reduction of bottlenecks. |
| Operational cost reduction | Smaller operational and data management costs of local devices versus cloud systems and data centre costs. |
| Security | End-user has the ownership of data and increased privacy is gained by reducing un-necessary data transmitting to cloud. |

Table 8. Edge challenges and their descriptions

| **Challenge** | **Description** |
|---|---|
| Availability and reliability | Improvement on these is established with redundancy which results in energy consumption hike. |
| Capacity of storage | Edge (IoT) devices are not suitable for storing large amounts of data. Edge utilizes distributed data storage method and Edge layer has bigger, but limited storage resources compared to device and cloud layers. |

| Challenge | Description |
|---|---|
| Heterogeneous platform | Implementation of applications is hard due to heterogeneous platforms and devices. This also makes management of data and resources hard. |
| Initial & maintenance cost | The costs related to edge computing are high at the beginning when investment to the HW is done. In addition to this, also maintenance costs need through the life cycle of the system to be considered along with SW related cost e.g., SW infrastructure updates which require resources. |
| Network complexity | When edge system becomes large and heterogenous, the complexity of it makes it difficult and expensive to maintain. |
| Processing power | Limited processing power at the edge (IoT) devices. Complex calculations and large amount of data require lots of processing power. |

Bhattacharya (*2022*) discusses about advantages of edge computing in pharma industry and lists these in a structured way. Compliance is one benefit that he highlights, and it is very relevant to the working environment of this thesis. As he mentions IoT products are able to provide audit trails and data integrity which is very significant requirement in pharmaceutical industry (see Chapter 2.1). He also concludes that edge technology enables improvements to batch cycles and digital compliance, and underlines business perspective on how adaption of edge will evolve by mentioning Return On Investment (ROI) expectations that drive business decisions.

### 2.3.4 Architecture

Edge computing has many different architectures that companies can implement based on their needs, existing applications and future plans see e.g., (*Edge Computing Reference Architecture 2.0, n.d.; FAR-EDGE Reference Architecture – Edge4Industry, 2018; Lin & Simmon, 2022*). Simplified edge architecture is presented in Figure 6. This three-layered architecture consists of device layer, that is the source of data and where simple processing can be performed, edge layer, where near real-time processing happens and cloud layer which stores the data and performs the not real-time processing. (*Pérez et al., 2022; "What Is Edge Computing," 2025.*)

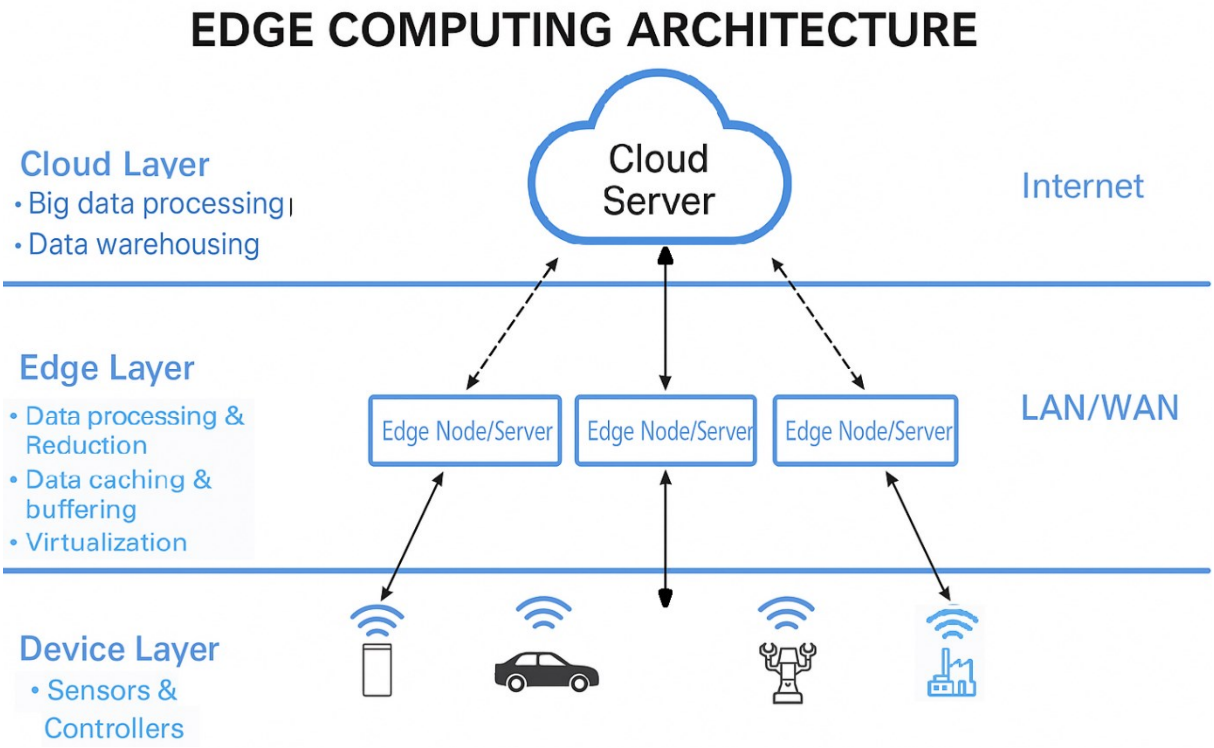



Figure 6. Edge architecture simplified (Adapted from *“What Is Edge Computing?,” 2025*).[1]

In Figure 7, the reference architecture for edge computing is shown. This architecture is designed by Edge Computing Consortium (ECC) and Alliance of industrial Internet (AII) to attain coordination between the digital and physical world. It tries to achieve cross-industry collaboration and reduce system heterogeneity. Its goal is to enable simplified cross-platform migration, and to effectively give support to system lifecycle activities. At horizontal perspective, it provides model-based open interfaces and at vertical perspective, architecture delivers smart services using management services, lifecycle data services, and security services to the entire service process and full lifecycle. (*Edge Computing Reference Architecture 2.0, n.d., pp. 15–17; What Is Reference Architecture? | Hpe.Com, n.d.*.)

[1] Image has been created using the help of AI (OpenAI’s Sora)

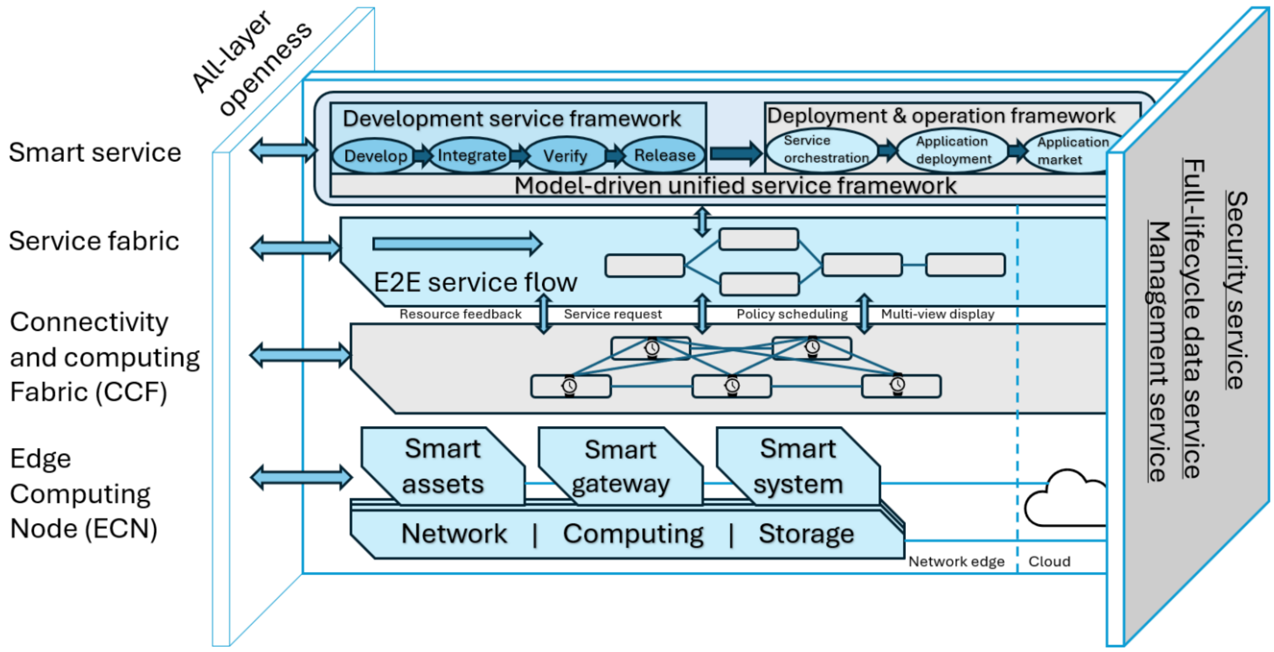



Figure 7. Edge computing reference architecture (*Edge Computing Reference Architecture 2.0, 2017, p. 16*).

To define the layers in Figure 7, ECC and AII have written unambiguous characteristics of each as follows:

- **Smart service** layer is model-driven unified service framework. Consistent SW development interfaces, and automatic deployment & operations are enabled with intelligent coordination between deployment service and deployment & operation frameworks.

- **Service fabric** defines E2E service flow by orchestrating smart services to realize service agility.

- **Connectivity and Computing Fabric** enables architecture to be more simplified, and it simplifies the distributed edge intelligence architecture for services. By supporting coordination between service needs of industries and edge computing resource services, It enables automatic and visualized deployment and operations of the Operational Technology (OT) and ICT infrastructure.

- **Edge Computing Network** delivers intelligent HW and SW security, it supports real-time response and processing, and it is compatible with a range of heterogenous connections.

Edge computing reference architecture and reference architecture in general is a document or a collection of documents that are standardized framework and/or blueprint. These frameworks outlie recommended structures and integrations of IT services and products to deliver a solution. They consist of industry best practises, and they typically promote most effective way to implement technologies and systems in a format that is easy-to-understand. Reference architectures are designed to help project managers and IT managers as well as developers, enterprise architects and other team members to collaborate and communicate effectively with each other regarding implementation project. (*What Is Reference Architecture? | hpe.com, n.d.*.)

## 2.4 Fog computing

Fog computing is "new" computing paradigm, which has its origins dating back to early-mid 2010's when Cisco introduced the term and later OpenFog consortium was formed by academical institutions, government organizations and industry leaders (*Kaswan et al., 2024, pp. 1–2*). It is a distributed, decentralized architecture and it can provide services in the same way as cloud does, difference is that these services are near the IoT and edge devices (*Pushpa & Kalyani, 2020*). By NIST definition, fog computing is layered model that enables ubiquitous access to shared and scalable resources (*Iorga et al., 2018*). Fog computing connects centralized cloud and distributed edge devices, and it provides new architecture that has critical role in coordinating and managing the network and resources to meet the quickly changing service requirements. Fog, together with cloud and edge, supports intelligent applications and sophisticated services such as AI and ML applications. Making computing resources more accessible, fog addresses infrastructure challenges and constrains associated with individual use of either cloud or edge computing, such as bandwidth, connectivity, latency, efficiency and cost-effectiveness. (*Yang et al., 2020, p. 1.*)

As seen in Figure 8, the three computing paradigms are tightly connected to each other and from the edge devices the data flows through fog and from there to the cloud. Fog being hybrid solution of cloud and edge extending cloud closer to the source of data giving the solution better latency and higher computational power compared to just cloud solution. (*Kazeem, 2018.*)

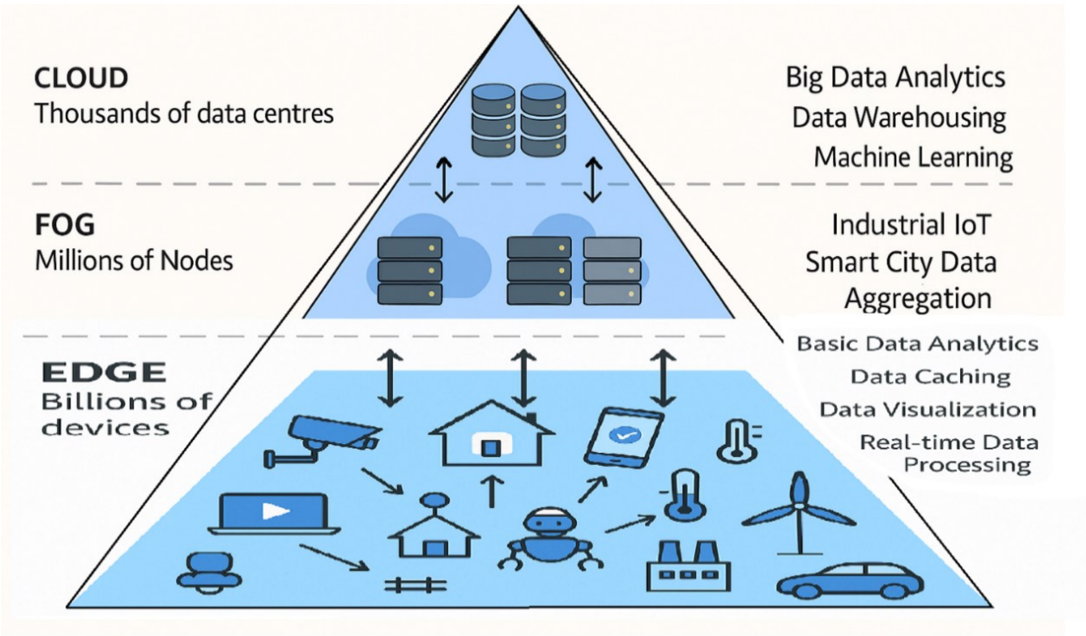



Figure 8. Fog architecture simplified (Adapted from *Kazeem, 2018*).[1]

Idea of the fog architecture is that not every data package is sent to the cloud to optimize cloud usage which affects the Operating Expense (OpEx) spending and to minimize the possible problems that come from latency dependent applications and computation. At its simplest way fog architecture is represented in 3-tier basis as seen in Figure 9, where the tier 1 is the end point data from e.g., sensor data from IoT devices. This is the raw data coming directly from the source. Tier 2 is the fog computing layer, “the fog/edge intelligence” as it is sometimes referred. This layer temporarily stores, computes and processes the data coming from tier 1. Tier 2 devices are mainly routers, switches and gateways, and also other devices that are capable and have hardware to perform the previously listed tasks. After the fog computing layer is done with the data, it sends the processed data is sent to the cloud framework. Tier 3 is the cloud computing layer. This layer is capable of storing and processing tremendous amount of data and there are wide set of applications available that the service providers have available. These are of course costly and therefore fog computing layer plays key role in the middle to reduce the costs of cloud computing. (*Taneja & Davy, 2016, pp. 154–155.*)

[1] Image has been created using the help of AI (OpenAI’s Sora).

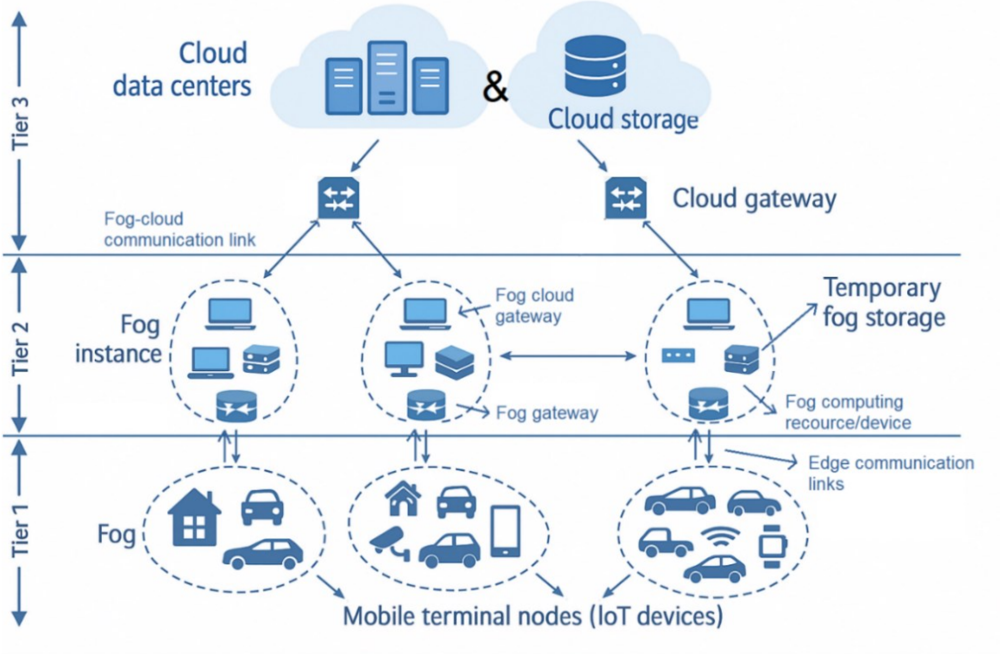



Figure 9. 3-Tier architecture of fog (Adapted from *Taneja & Davy, 2016*).[1]

Figure 10 represents the fog architecture with the highlighting the centred fog "nodes" and the gateways between the IoT and the fog nodes. In fog computing, the actual manipulation of data and the computation is done with proximity, meaning that it is done close to the nodes, the data source. Benefit of the proximity is network efficiency. With this, the benefits of classical architecture, edge architecture, are utilized and the limitations are avoided. (*Sandoval, 2017.*)

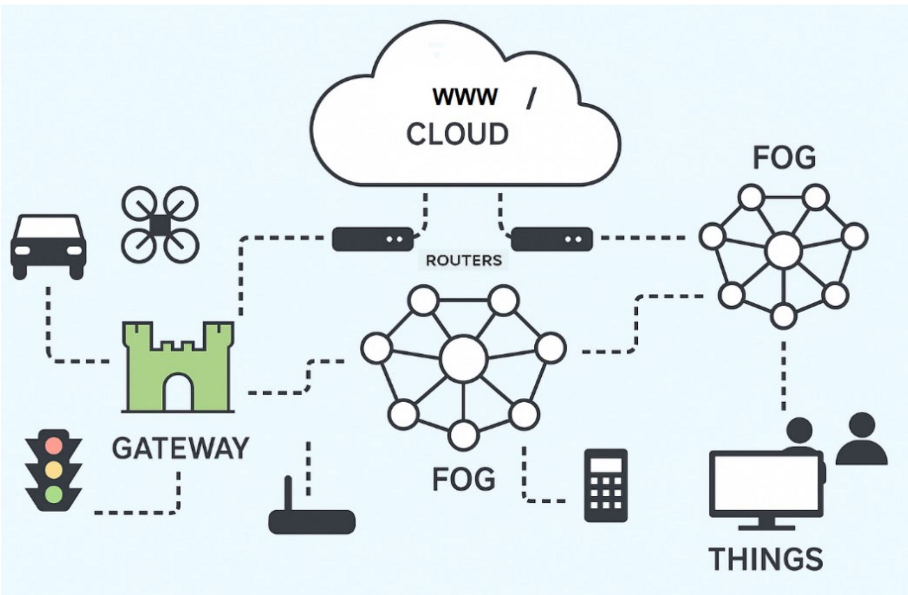



Figure 10. Fog nodes and gateways (Adapted from *Sandoval, 2017*).[1]

[1] Image has been created using the help of AI (OpenAI's Sora).

According to multiple research articles and papers such as (*Fog Computing in the Internet of Things: Intelligence at the Edge | SpringerLink, n.d.; Iorga et al., 2018; Kazeem, 2018; Pushpa & Kalyani, 2020; Sandoval, 2017; Taneja & Davy, 2016; Yi et al., 2015*), the benefits of fog are obvious and notable. These benefits include such as response time improvement and network efficiency due to proximity,  reduction of congestion (occurs when there is high traffic demand and the network can't handle that and data transfers get delayed or even lost), reduction of bottlenecks common in traditional architectures by de-centralizing network and spreading responsibilities across multiple nodes, nodes bring also scaling abilities to the system and the design increases the security of the system. (*Sandoval, 2017.*) Pushpa and Kalyan (*2020*) also list the following network parameters as benefits of edge/fog: speed, bandwidth, response time, latency and throughput. According to NIST publication of Fog computing Conceptual Model (*2018*) the essential characteristics of fog are for example contextual location awareness, fog-node clusters, heterogeneity, low latency, interoperability and federation, real-time interactions and scalability. The same characteristics and benefits appear in the literature and there are also examples of potential use cases where fog is the trivial paradigm to go with. To name a few, Sandoval (*2017*) presents e.g., A.I., multimedia rendering, Swarm computing and encoding as such applications where fog has major use case and the applications benefit hugely from it. Other applications where fog is important are healthcare applications which are sensitive to response time, and which are in need of distributed analytics (*Rahmani et al., 2018, p. 165*).

The disadvantages that are related to fog are limited resources, complex architecture and limited coverage. Use of fog computing when it is not really needed may mean that the architecture performs wors than it would without it. Fog ads lots of points of failure to the system and the maintenance and predictability suffer. Spreading out effort and responsibility means that there is loss of privacy in the core concept of fog and this needs to be mitigated if fog is used. (*Clancy, 2023; Sandoval, 2017.*) Yi et al. (*2015*) discuss about the design of fog computing platform and the challenges associated with it. According to them, the design of the platform is non-trivial if latency, efficiency and generality goals need to be met. They list four (4) challenges that they found, choice of virtualization technology, fight with latency, network management and security and privacy. These can also be seen as disadvantages of fog, if the architecture and platform design is not well designed.

### 2.5 Large Language Models

LLMs are gaining a lot of attention in the public and private sectors. The legislation around AI has been in the centre of discussion as the new AI act entered into force on 1st of August 2024. This regulation lays down harmonized and risk-based rules on AI, and it is first comprehensive legal framework on AI in the world. It defines the level of risk regarding AI to 4 levels: Minimal risk, Limited risk, High risk and Unacceptable risk. These levels determine what rules apply on each AI system, are they banned immediately or if they have no rules applied to them at all. (*AI Act | Shaping Europe's Digital Future, 2025.*) It also takes into consideration LLMs as they are general-purpose AI models and as AI act gives example of this: "Large generative AI models are a typical example for a general-purpose AI model, given that they allow for flexible generation of content, such as in the form of text, audio, images or video, that can readily accommodate a wide range of distinctive tasks." (*REGULATION (EU) 2024/1689, 2024, p. 26*). Generative AI such as ChatGPT, are not classified as high-risk regarding the levels AI Act defines (*EU AI Act, 2023*).

LLMs are massive deep learning models, and they trained with massive amount of data which allows them to understand and generate natural language. They are also capable of performing wide range of different tasks including image, text and voice generation. The capability of Natural Language Understanding & Processing (NLU & NLP) has drastically increased over time as IBM, Google and other big players have focused on this technology alongside with e.g. ML, Neural Networks (NN) and transformer models. LLMs are designed to generate text and other content, and they are able to generate coherent and contextual text with multiple languages. The underlying transformer in LLMs is NN that is combination of coder and encoder. This transformer extracts meaning and context from the content input and is able to comprehend relations between phrases and individual words. (*What Are Large Language Models (LLMs)? | IBM, 2023; What Is LLM? | AWS, n.d.*.)

Transformers consist of multiple transformer blocks, layers. The layers can be for example normalization layers, feed-forward layers and self-attention layers that are connected, and they interpret the input and generate the output. Two main reasons why LLMs have adopted transformers are positional encoding and self-attention. Positional encoder enables the transformer to understand the order of words in a sequence as the transformer processes word in parallel and therefore does not know the order of them. Self-attention allows the model to assign weights to the individual parts of the input to let it focus on the important elements of the content. (*What Are Large Language Models? | NVIDIA Glossary, n.d.*.)

Figure 11 is example architecture of transformer stack. In this example the encoder takes an input sequence of symbol representations ($x_1$, ..., $x_n$) and transforms it into a sequence of continuous vectors z = ($z_1$, ..., $z_n$). Using this encoded representation z, the decoder then produces an output sequence ($y_1$, ..., $y_m$), generating one symbol at a time. This process is auto regressive, meaning the decoder uses the symbols it has already generated as part of the input for generating the next one. The Transformer model adopts this general structure by using layers of self-attention and fully connected feed-forward networks in both the encoder and decoder. (*Vaswani et al., 2017, pp. 2–3.*)

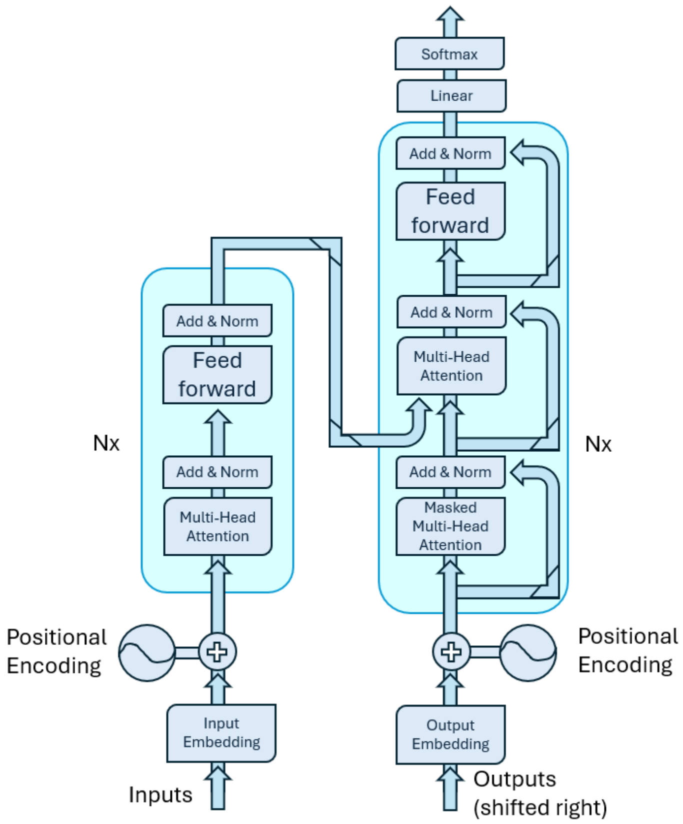



Figure 11. Example transformer architecture (adapted from *Vaswani et al., 2017, p. 3*)

Large language models as decision-making tool appear in literature quite many times (*Eigner & Händler, 2024; Elin, 2023; Gumilar et al., 2024; Meccyjoy et al., 2024; Wang et al., 2025*). Some of these are from the medical/pharma field and others are more universal. Understanding the possibilities and limitations of using LLMs in the decision-making process allows utilizing modern technology to help minimize time consumption and make generic decisions automated. Wang et al. propose human-expert-level framework to handle multi-criteria decision-making tasks with appropriate prompt design and engineering. They give valuable insight into the development of intelligent decision-making systems that utilize LLMs. Meccyjoy et al. (*2024*) discuss LLMs being pivotal tools that are transforming decision-making. They explore the impact that LLMs are having on shaping smart, adaptive and data-driven decision-making. This is remarkably significant regarding this thesis. The things that LLMs and AI-augmented future have to offer are beyond today's human biased, inefficient decision-making processes with innovative perspectives and strategic foresights. Meccyjoy et al. (*2024*) go deeper into real-world integration and human-AI collaboration model which fits perfectly to the scope of the thesis. They see LLMs having transformative impact across various domains from finance to healthcare. They point out advantages of using LLMs for decision-making, practical challenges related to it and the future of LLMs in decision-making.

Advantages that they mention are, for example are previously mentioned minimization of human bias, data-driven objectivity, and speed and scalability. Practical challenges that they mention are e.g., Bias in AI models, Over dependence on AI and Continuous learning and adaptability of models. And for the future they think that LLMs will have evolving roles in leadership and governance, grooving collaboration of humans and AI, and next-gen LLMs evolving to continuously leaning and adapting models. (*Meccyjoy et al., 2024.*)

Other interesting and relevant study has been conducted by Elin, who focus on potential use of AI in decision support systems. He compares the ability of two LLMs compared to small sample size of human participants interpretation and definition of objectivity. Results that he was able to get suggest that LLMs need customization for specific application or task to get more accurate outputs and to enhance the decision-making abilities of them. This study was conducted with two popular LLM models and with 86 human participants. The LLMs answers were 75% agreeing with human responses and comparing the human responses with each other, there was 100% agreement. This highlights the need for customization of the LLMs and the importance of continuous monitoring by humans to assure successful decision-making. (*Elin, 2023.*)

### 2.6 Literature summary

The aim of the literature review was to familiarize reader to the operating environment of pharma, the regulatory entities and the regulation associated with computerized systems, the paradigms that are being evaluated in this thesis, and quickly go through fog computing which appeared frequently in the literature when cloud and edge computing paradigms were discussed. Also, LLMs are briefly presented in a sense of being a helping tool for decision-making process.

Very little research was found that related directly to the comparison of cloud and edge solutions in a pharma environment. Hence the literature review is constructed of each topic being presented separately and the connection between each is implemented in the upcoming chapters. As a result of the literature review, a survey is being designed to answer the first research question.

The pharmaceutical operating environment presents vast amount of regulation and regulatory authorities both globally and nationally which are crucial in the pharmaceutical manufacturing to ensure patient safety and product quality. The regulatory authorities such as FDA and Fimea ensure that companies producing in human and/or veterinary medicinal products comply with the regulations and guidance.

Guidance that gives the pharma companies the framework in which to operate, comes from e.g., GMP and GAMP 5. These are industry best practises and by following them manufacturers ensure compliance of regulations. Computerized systems for example, are governed in Annex 11 which gives guidance for the validation and other aspects of CS lifecycle.

Neologisms Industry 4.0, Pharma 4.0 and RAMI 4.0 are part of global technological transformation toward smart manufacturing. These are changing the manufacturing industry and pharma is part of this change, applying 4IR innovations and concepts. In pharma industry, the adaption and change are slower compared to other manufacturing industries, and this is partly the result of the high regulation. The three-dimensional reference architecture of RAMI 4.0 is part of 4IR, and it is supporting its initiatives by ensuring common understanding of life-cycle management, enterprise-control system integration and batch control. It can be used to develop future business models and products in pharma industry.

Based on the literature reviewed, cloud computing offers multiple benefits for companies to utilize and implement, that increase their proficiency and profitability. Cloud computing as a paradigm, is not so new, and the story dates back to as early as mid-1940’s. From the early days, cloud computing has evolved into multibillion-dollar business and large tech companies dominate the field.

Cloud computing offers users different service and deployment models from which to choose the most desirable option. The most know service models are SaaS, PaaS and IaaS. These are the standard when implementation of cloud is considered but also newer once exist, for example AIaaS, CaaS and FaaS are also products that cloud providers are offering for customers today. The deployment models are also very important to be considered when designing cloud implementation. Options are private cloud, community cloud, public cloud and hybrid cloud. These models affect e.g., to the level of privacy of the data. Private cloud, as its name suggests, is exclusively dedicated to the use of individual organization and public cloud is used open for multiple users and organizations at the same time. Hybrid cloud combines these two and it also utilizes on-prem infrastructure which helps to tackle fundamental issues of cloud computing such as latency problems.

Architectural options that cloud computing provides vary a lot and based on the application and solutions needed, the architecture can chance. In Figure 12, simplification of cloud computing paradigm is presented. More detailed and standardized architecture the NIST reference architecture where five major actors cloud auditor, broker, carrier, consumer and provider are presented this can be seen in Figure 5. This is a high-level presentation of architecture, but it gives the viewer understanding of the characteristics, requirements, standards and uses of cloud computing.

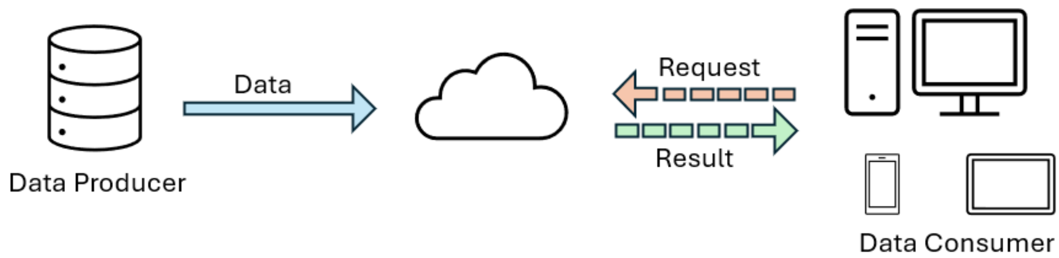



Figure 12. Cloud computing paradigm (adapted form *Shi et al., 2016, p. 638*).

Most important take in Chapter 2.2 is the benefits and limitations of the cloud. These are listed clearly in the Table 5 and Table 6. In the literature, these benefits and limitations were presented differently depending on the contexts and the viewing perspective of given source reviewed and these two tables are formed deductively based on the factors affecting this thesis, the research questions and the operating environment at hand.

Edge computing offers computing services at close proximity of the data source, and it offers solutions to cloud drawbacks e.g., latency and bandwidth. The term "edge computing" originates back to early 2000 when it was firstly introduced. From there the paradigm has evolved and gained wide recognition in industry. With the help of IoT device development and adaption, edge computing has become possible. These devices have capabilities to process the data before sending it to the edge nodes which reduces the HW requirements of the nodes.

Visualization of edge computing paradigm can be seen in Figure 13. The architecture of edge computing is three-layer architecture where the edge layer is located in between cloud layer and device layer this is shown in Figure 7. There is also a reference architecture for edge computing which is standardized framework that outlies the recommended structures and integrations of IT services and products.

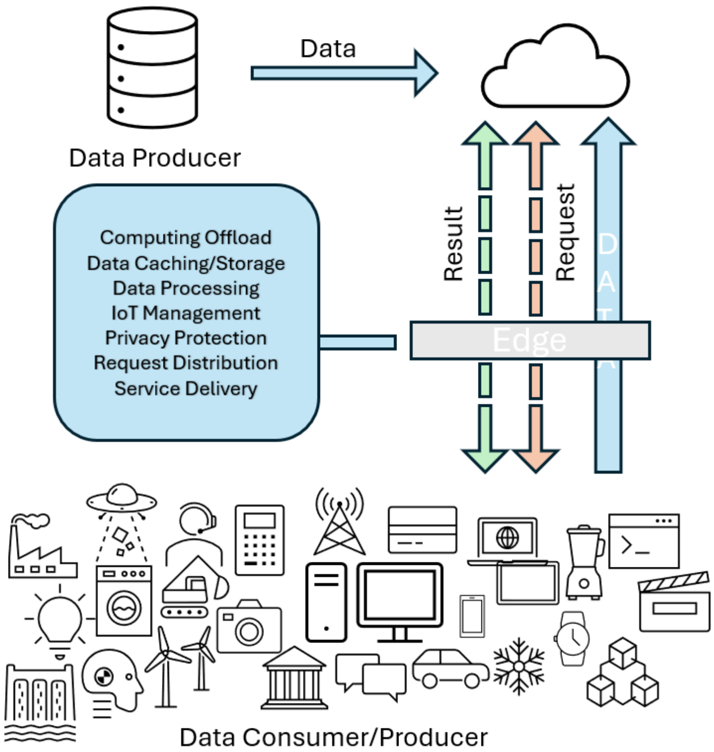



Figure 13. Edge computing paradigm (adapted from *Shi et al., 2016, p. 638*).

Key takes of the Chapter 2.3 are the benefits of edge come from the ability to process data at the edge of the network, meaning that its done close to the actual source of the data which eliminates the time consumed to transfer the data to off-prem cloud. Other benefits and limitations of edge computing are described in Table 7 and Table 8. These characteristics are deductively used to form the survey that is presented in Chapter 3.

Fog computing connects distributed edge devices and centralized cloud. Its role is to manage and coordinate resources and the network. It can be used in intelligent applications such as ML and AI applications together with cloud and edge. It is three-layer architecture, and the fog nodes are located between cloud and edge nodes. Key benefit of fog is the cloud optimization it is capable of doing by processing data before cloud resulting in OpEx savings. Fog computing paradigm is shown in Figure 14.

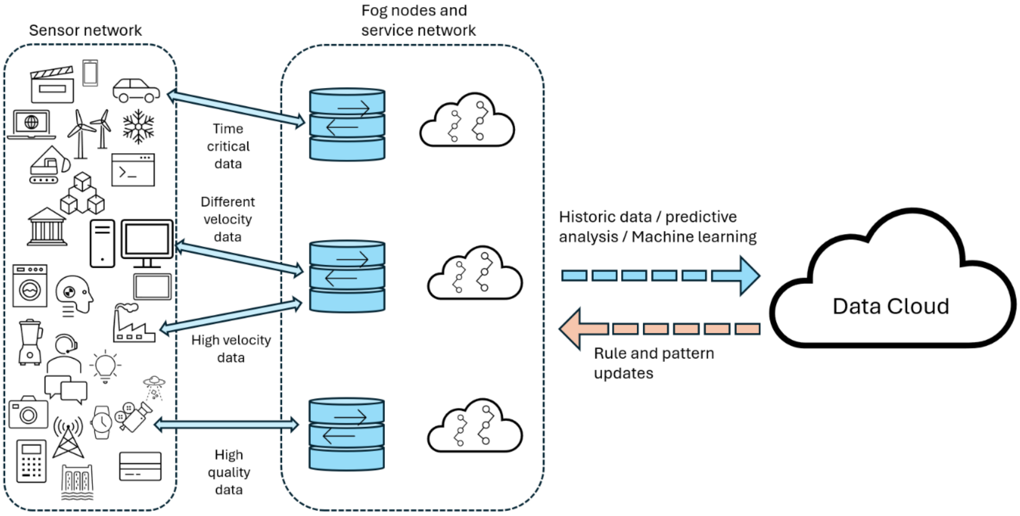



Figure 14. Fog computing paradigm (Adapted from *Fog Computing and Internet of Things | Techplayon, 2017*).

To be able to perceive the tie-up of cloud, edge, fog and IoT Figure 15 shows the whole computing stack and Table 9 summarizes cloud, edge, fog and IoT characteristics and it allows easy comparison between all four.

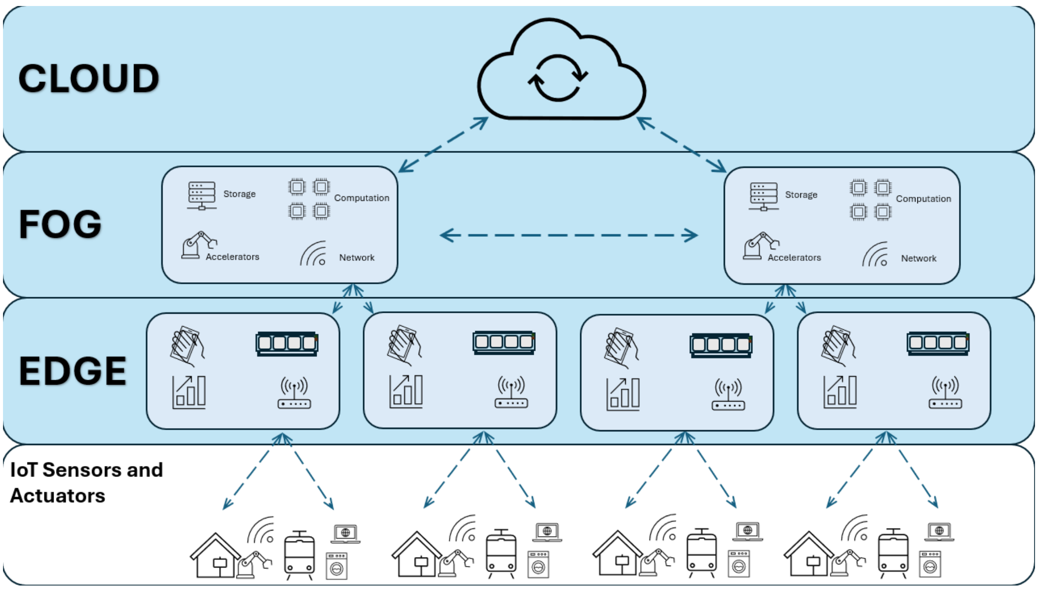



Figure 15. Whole computing stack (adapted from *Maciel et al., 2022, p. 235*).

Table 9. Characteristics of each paradigm (adapted from *Taheri & Deng, 2020, p. 10*).

| Characteristic | Cloud | Edge | Fog | IoT |
|---|---|---|---|---|
| Components | Virtual resources | Edge nodes | Fog nodes | Physical devices |
| Computational | Unlimited | Limited | Limited | Limited |
| Data | Process | Process | Process | Source |
| Deployment | Centralized | Distributed | Distributed | Distributed |
| Distance to data source | Far | The nearest | Near | Source |
| Location awareness | Not aware | Aware | Aware | Aware |
| Nodes count | Small | Very large | Large | The largest |
| Response time | Slow | The fastest | Fast | - |
| Storage | Unlimited | Limited | Limited | Very limited |

LLMs introduce a new way of automating tasks and decision making. Especially generative AI has been hyped, and it has gained lots of media attention in the past few years as models such as ChatGPT have revolutionized the field. Quality of text, image and sound generation of LLM models has increased dramatically over the years and these models can be used to benefit and streamline business processes. In pharmaceutical industry the regulation has already picked up on this new trend and for example the new version on GAMP 5 deals with the aspects of using AI tools in the GxP environment. Key point of Chapter 2.5 is the example of the use of LLMs in decision-making process and the accuracy it is able to provide when the customization of the model is done correctly.

# 3 Survey and Interview

As part of the Mixed Methods Research (MMR) – a combination of quantitative and qualitative research methods – there are many different designs that exist, for example convergent mixed methods, explanatory sequential mixed methods and exploratory sequential mixed methods. This thesis utilizes the explanatory sequential mixed methods design, where the research first focuses on the quantitative data and its analysis and then the qualitative research, data and analysis follows. (*J. W. Creswell, 2018, p. 13.*)

As a quantitative research method, a survey is held to a carefully picked focus group. This group of people consists of IT professionals, technical professionals and otherwise technologically oriented people. Also process and business-oriented people and validation experts were included. To gain broader understanding and business focused insight from the results of the survey, a semi-structured interviews are held after participants have answered to the survey. The purpose of these semi-structured interviews is to get qualitative data from the participants that would not come up in the survey alone. These two combined, help build the most useful decision matrix and guideline for the target company as deductively compiled survey is enriched with inductively interpreted interview.

This chapter answers the first research question by mapping the critical characteristics that allow the comparison of the computational solutions in a pharma related environment, and which are affecting the decision-making process.

## 3.1 Survey design and focus group

The survey questions are built to collect understanding of the importance of individual parameters and e.g., the relevancy of the regulated environment and desirable expense structure of system when evaluating the appropriate implementation method in individual cases regarding Cloud and Edge solutions. The result of the survey is used to help design the decision guideline and to weight the parameters in the decision matrix. The survey was originally planned to be sent to the focus group as a preliminary information that they could view before the interview, but it was noted that it would be more efficient to turn the tables and conduct a survey and have the interview only as a summary of the survey. This way the focus group is able to rate the importance of each parameter and question in a scale of one (1) to four (4), one being “not important” and 4 being "very important”. These results are used in combination with the inductively interpreted interviews to give weight to the parameters in the decision matrix presented in Table 13.

The list of questions was generated in Microsoft Forms to allow respondents to evaluate the importance of the individual parameters as they see them. Final results of the questions are represented in Forms own summary of results that can be seen in Appendix 2 on page 96.

The focus group (respondents) are selected based on them being a part of the innovation venture consortium, and the projects that are related to it and ongoing at the company. Focus group has representative of technical, IT, and validation professionals, process people, businesspeople and project managers. Aim of this distribution is to reduce the possibility of biased results in the survey by only concluding e.g. technical professionals and not considering the possibility of business point of view. The surveys are sent to the respondents via email link which directs them to the Microsoft Forms survey. After limited response time, each person that received the survey are asked to have either via Teams or face-to-face meetings in order to go through their answers verbally to get more insight of their thoughts on each question. These meetings are recorded to allow more thorough analysis of the conversation. This is of course agreed with each respondent separately before the interview starts. The recordings are deleted after the analysis is completed and there is no further need for the recordings. To prevent any GDPR violations, the respondents are labelled as "Respondent 1, 2, 3, etc." and no names are collected or used in this thesis. Forms has the option to collect participants name, but for this survey the option "Anyone with the link can respond" was used and for this option, the "Anonymous answer, no login required" apply. Table 10 shows the response rate and whether results were gone through verbally with the respondent afterwards.

Table 10. Response rate and interviews held.

| Respondent | Survey completed | Interviewed |
|---|---|---|
| Respondent 1 | ✓ | ✓ |
| Respondent 2 | ✓ | ✓ |
| Respondent 3 | ✓ | ✓ |
| Respondent 4 | ✓ | ✓ |
| Respondent 5 | ✓ | ✓ |
| Respondent 6 | ✓ | ✓ |
| Respondent 7 | ✓ | ✓ |
| Respondent 8 | ✓ | |

| Respondent | Survey completed | Interviewed |
| --- | --- | --- |
| Respondent 9 | ✕ | |
| Respondent 10 | ✕ | |
| Respondent 11 | ✕ | |

### 3.2 Results and conclusions

The response rate was expected, and the quality of the responses was good. Forms gives clear summary of results which can be seen in Appendix 3 on page 100. In addition to this, the answers were shifted to Excel to further analysis. These graphs can be seen below in Figure 16, Figure 17, Figure 18 and Figure 19.

Figure 16 and Figure 18 represent the results of the individual questions and show the variance between respondents. From these, the distribution of answers can be seen. Figure 17 and Figure 19 are calculated values of the results showing the min and max values of the rated questions and the average of these. Also, the deviation of answers is represented in these two images with yellow line and the three black dots in each of the images highlight the questions with the largest deviations in respondents' ratings. Deviation function that was used in Excel was STDEV.S() which is calculated as the standard deviation of sample with the following formula:

$$\sqrt{\frac{\sum(x-\overline{x})^2}{(n-1)}}$$

Where,

$x$ = the value in the data distribution,
$\overline{x}$ = $the\ sample\ mean$ and
$n$ = total number of observations.

This function measures how widely values are dispersed from the average value (mean value). For the sake of clarity deviation was selected here to better and more intuitively represent the spread across respondents' answers to the questions. If respondent answers "No" or "Maybe" to the questions 2-9, they are not prompted with the follow-up question asking them to rate the importance of the given question at hand and this will show as 0 in the Figure 18. In the spreadsheet these cells are "FALSE", and this value is not included in the average or deviation calculation.

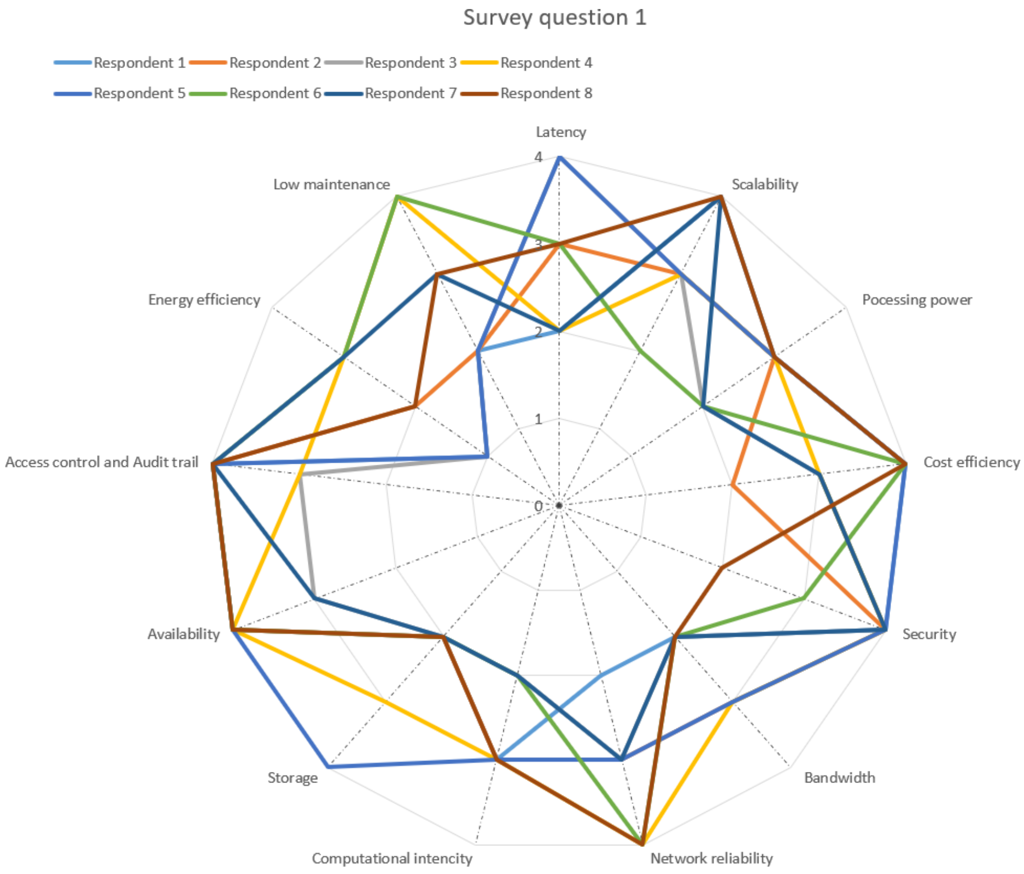



Figure 16. Radar chart of question 1 answers.

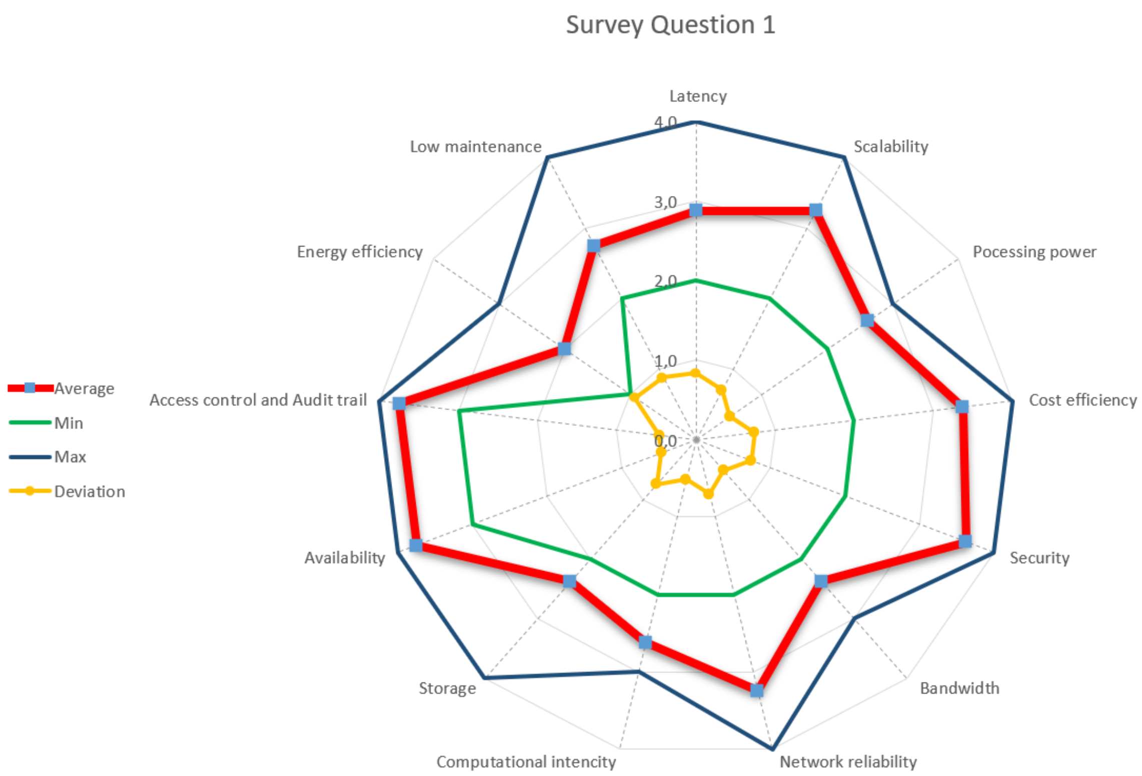



Figure 17. Calculated values of question 1 answers.

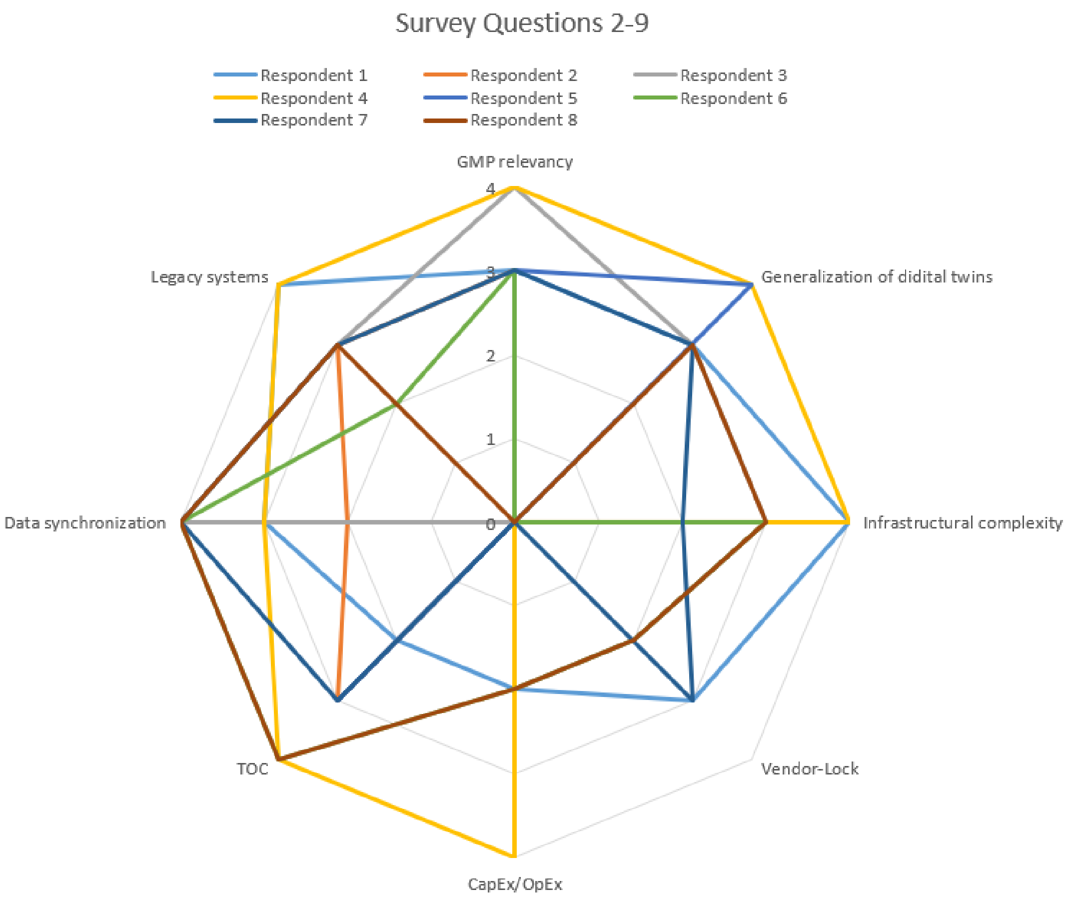



Figure 18. Radar chart of question 2-9 answers.


Figure 19. Calculated values of question 2-9 answers.

Figure 17 and Figure 19 were the most helpful regarding the design of the decision matrix. From the values show in these, the calculated deviation values were used to determine the largest difference in respondents´ answers  and this shows the heterogeneity in the respondents' field of expertise which is very interesting to see in the results. It also shows that respondents have interpreted the questions in different way, and this was also noted in the semi-structured interviews when each respondent was able to tell their thoughts on the survey. The calculated average values were used to sort the characteristics in the decision matrix so that the highest average values were weighted to be most critical in the decision-making process and the lowest average values were weighted to be least critical.

To get more comprehensive conclusions of the whole survey, the respondents were asked to have short teams meeting to go through their answers verbally to get insights into their thoughts on individual questions and to possibly get relevant information about their answers that cannot be expressed via the survey. These semi-structured interviews helped to inductively give weights to the characteristics in the decision matrix presented in Chapter 4.1.

In the interviews, the respondents were told the goals that the survey and the interview had, and they were told that the decision guideline was the desired throughput of the entire thesis work. Based on this, they were able to answer better to the individual ratings they had given to the survey questions, and they were able to tell their thoughts and opinions regarding the whole survey and individual questions in it. The benefit of having the interview held, was to get broader understanding of the results that the survey had, and it also highlighted individual characteristic better than the survey could have highlighted individually.

# 4 Building guideline for decision making

In this chapter, the guideline for decision-making and the decision matrix is presented starting with the first version of the matrix with brief explanation on how its designed and how it is working. Next, the second version is introduced and the modifications to the prior version are explained. After this, the third and the final version is presented with the inductively weighted values gained from the combination of literature revive, survey and interviews held with the focus croup. Then the guideline for decision-making is gone through and its value for the company is demonstrated with three use-case examples.

This chapter answers the second research question, and it gives the reader understanding on the importance of the decision matrix in the decision-making process in regulated pharma environment, how important it is to have standardised process on how to evaluate the needed/wanted computational solution for each project, case and system.

## 4.1 Contents of decision-making guideline and decision matrix

First version of the decision matrix shown in Table 11 is quite simple and it only shows individual characteristics as determinant factors for each option. If “Latency” is the key characteristic that defines which option is suitable it shows in the matrix that Edge is the better solution. But if both “Latency” and “Scalability” are key characteristics there is no clear solution on which implementation should be used. There is third solution “Fog” that gives “Possible with trade-offs” option in this case but it is not very helpful or straight forward for the person evaluating the use-case. The main reason why Fog computing has been added to this decision matrix is that it illustrates the possibility of third option that is somewhere in the middle of Edge and Cloud, and it is a compromise of the two. Fog brings value to the organization needing solution regarding cloud versus edge by making the decision easier, but it has trade-offs as mentioned. The matrix does not consider the variance and the amount of importance of each characteristic. For example, “Network reliability required” is not very good factor regarding the decision-making if the importance of it is just “Yes” or “No”, it will give much more value to the comparison when it can be rated from 1 to 4 (1 = Not important and 5 = Very important). This way the combination of characteristics can be taken into consideration and weighted sum can be calculated which then favour on solution over the other.

Table 11. 1st version of the Decision matrix.

| Characteristic | Cloud | Edge | Fog |
|---|---|---|---|
| Latency sensitivity | ✕ Low | ✓ High | ⚠ Medium |
| Compute intensity | ✓ High | ✕ Low | ✓ Medium |
| Local uptime requirement | ✕ | ✓ | ✓ |
| Regulatory data locality | ✕ | ✓ | ✓ |
| Scalability needs | ✓ | ✕ | ⚠ |
| Network reliability required | ✓ | ✕ | ⚠ |
| Security (Physical exposure) | ✓ | ✕ | ⚠ |
| Budget model (OpEx focus) | ✓ | ✕ | ⚠ |
| Data aggregation needs | ✕ | ✕ | ✓ |
| Maintenance burden | ✓ Low | ✕ High | ⚠ Medium |

✓ Best fit ⚠ Possible w/ trade-offs ✕ Poor fit

The second version of the Decision matrix shown in Table 12, includes the scale from 1 to 4 which illustrates the importance of each characteristic involved in the decision-making. This allows better understanding of individual needs and drivers when the guideline is being used to evaluate the go-to option for given use-case. From this version, the column "Fog" has been deleted to emphasize the 2nd research question presented in Chapter 1.

Table 12. 2nd version of the Decision matrix.

| Characteristic | Importance (%) | USER defined importance (scale 1-4) | Cloud weighted score | Edge weighted score |
|---|---|---|---|---|
| Cost | 0,30 | 1 | 0,198 | 0,099 |
| Real-time Processing | 0,25 | 3 | 0,2475 | 0,495 |
| Latency Sensitivity | 0,20 | 3 | 0,198 | 0,396 |
| Data Privacy | 0,15 | 4 | 0,198 | 0,396 |
| Scalability | 0,10 | 2 | 0,132 | 0,066 |
| Total | 100 % | | 0,9735 | 1,452 |

| Fixed values | User defined |
|---|---|
| | |

| Suggested computational implementation to go with |
|---|
| **EDGE** |

Final decision matrix can be seen in Table 13. In this version, all of the characteristics that are considered important are added to the table and the "Weight of the characteristic" value is fixed for each. These fixed values are the result of the survey, and the discussions held with each respondent and based on this the weighted values are inductively given to each characteristic. For example, the Total Cost of ownership contains both Capital Expenditure (CapEx) and OpEx, and the cost structure of the implementation and importance is not evaluated individually also vendor-lock has not been included in the final version of the matrix albeit being a part of the survey. The total sum of the characteristics equals to 100%. User defined importance scale is changed from earlier 1 - 4 to 0 – 4 (0 = Not Applicable (N/A), 1 = Not important and 5 = Very important). This allows the user to skip characteristics that are considered not relevant in the ongoing evaluation. Table 13 has values set to USER defined importance column for illustration of the functionality of the decision matrix. When the result of the matrix leans towards either, “CLOUD” or “EDGE”, it gives the suggested solution in the Suggested computational implementation to go with box, if no values have been set to the USER defined importance or if the sum of the Cloud and Edge weight score columns are close to equal, the matrix gives suggestion of “"CLOUD/EDGE”.

Table 13. 3rd version of the Decision matrix.

| Characteristic | Weigh of the characteristic (%) | USER defined importance (scale 0-4) | Cloud weighted score | Edge weighted score |
|---|---|---|---|---|
| Access Control & Audit Trail | 9,0 % | 0 | 0 | 0 |
| Availability | 9,0 % | 0 | 0 | 0 |
| Cost efficiency | 8,0 % | 0 | 0 | 0 |
| Data Privacy | 7,0 % | 0 | 0 | 0 |
| GMP Environment | 7,5 % | 0 | 0 | 0 |
| Infrastructure Complexity | 7,0 % | 0 | 0 | 0 |
| Latency Sensitivity | 4,0 % | 0 | 0 | 0 |
| Low Maintenance | 5,0 % | 0 | 0 | 0 |
| Network Reliability | 7,0 % | 0 | 0 | 0 |
| Processing Power | 3,5 % | 0 | 0 | 0 |
| Scalability | 8,5 % | 0 | 0 | 0 |
| Security | 9,5 % | 0 | 0 | 0 |
| Total Cost of Ownership | 15,0 % | 0 | 0 | 0 |
| **Total** | **100 %** | | **0** | **0** |

Fixed values/ Calculated values

User defined values

Suggested computational implementation to go with

**CLOUD/EDGE**

To get the matrix to give Suggested computational implementation to go with, each cell in Cloud weight score and Edge weight score has individual weights that determine each rows impact on the decision. For example, in row Latency the weights are set to emphasize Edge over Cloud and in the opposite  Low Maintenance emphasizes Cloud over Edge.

The “blue-pencil” decision-making guideline can be viewed in appendix 4 on page 102. This guideline has been stripped away of the company logos etc. that are considered as confidential. The purpose of this guideline is to help in the design phase when the evaluation is done whether to use on-prem (as in edge computing) or off-prem (as in cloud computing) solution for data storing and computation.

The guideline is “one-pager”, and it aims to guide the person making the decision to use the decision matrix to easily evaluate the criteria that the given system has for the computing solution. The guideline consists of short description on how to use the decision matrix and the decision matrix itself.

Considering the second research question presented in Chapter 1, it is possible and highly needed to create the decision-making guideline for regulated pharma environment. It serves as basis for the decision-making process, and it allows the person making the decisions to repeatably get same result with the same criteria and the user defined importance. It is also highly needful as it gives business value by affecting efficiency and consistency. This need and the value that the guideline creates, are also something that are recognized by the target company. Supporting view of this came up in a meeting held with one of the supervisors where they stated that: “I think that it would be very good to always include this decision-making guideline to the business case calculations and the evaluation of a new system introduction.” (*S. Gambier, personal communication, September 1, 2025*).

It also needs to be pointed out that even though the decision matrix gives unambiguous result on whether to go with cloud or edge solution, it is the individual’s own discretion whether to diverge from the solution proposed by the matrix. This type of situation might occur when the weights in “Importance” do not align with the persons own conception of the appropriate weights or they value individual character, such as the Total Cost of Ownership (TCO), over everything else.

## 4.2 Use-case examples evaluated with decision matrix

In this chapter, three real life use-cases from the target company are evaluated with the decision-making matrix. The aim here is to demonstrate the usability of the matrix and to make it easier for the reader to understand the real-life value that this master's thesis throughput is built to create and the and the applications it can be used in. This evaluation also brings valuable insight into the existing methods and processes that steer the decision in current projects and decisions. In each sub-chapter, one use-case is briefly gone through, and the current status of the implementation solution is mentioned and if the decision matrix suggested solution differs from this, the reasons are discussed. Each use-case is evaluated and rated by subject matter experts who are working in the given projects.

These use-cases have been selected based on their relevancy to this thesis. The use-cases represent three different levels of scale of applications. Starting from the largest implementation and then gradually moving to smaller implementation to see if the decision matrix is scalable and the use can be universal. To make the results of this chapter clearer, only the characteristic and user defined importance columns, and the resulting suggestion of the decision-making matrix (shown in Table 13) are presented.

### 4.2.1 MES

In the target company, corporation wide MES upgrade is being implemented and old – on-premises – solution is being upgraded to newer version that is running in private cloud. This large-scale application is in validation phase of the production environment, and it is yet to pass the strict requirements that the GMP sets to it. The evaluation of the cloud versus edge has been done and the implementation that has been selected is cloud.

Table 14 shows user defined importance values rated by subject matter expert of MES in the target company. He has been part of the MES project for a long time, and he has been doing similar evaluation internally in early stages of the project. This evaluation and the discussion of the current implementation have been done verbally via Teams. Each characteristic has been gone through with him and the discussion regarding his view of the user defined importance values has been done alongside and the explanation for each value is thought thoroughly. For example, the last row in the matrix, TCO, is evaluated to have low importance and this is based on the subject matter experts own conception of the current situation of the MES implementation projects cost basis and the corporations view on this matter.

Table 14. User defined importance – MES.

| **Characteristic** | **USER defined importance (scale 0-4)** |
|---|---|
| Access Control & Audit Trail | 3 |
| Availability | 4 |
| Cost efficiency | 1 |
| Data Privacy | 3 |
| GMP Environment | 4 |
| Infrastructure Complexity | 2 |
| Latency Sensitivity | 3 |
| Low Maintenance | 2 |
| Network Reliability | 3 |
| Processing Power | 2 |
| Scalability | 3 |
| Security | 4 |
| Total Cost of Ownership | 2 |
| **Suggestion** | **EDGE** |

Based on the decision matrix, the suggested implementation is edge computing. This differs from the implementation method chosen by the corporation. The reason for the difference comes from the decision makers views on what is the most important characteristics in this given application and also it can be noted that corporation politics have great influence on this sort of large scale and corporation wide decisions that the decision-making matrix does not take into consideration. Worth mentioning is also the effect of GMP in this type of decision and implementation. The regulation related to validation of computerized systems and the data integrity (gone through in Chapter 2.1) make it hard to validate the system in pharmaceutical environment and to get it to production use, which is noticeable internally as the cloud implementation validation takes large amount of time and resources.

### 4.2.2 Track and trace

Track and trace system is used in the target company at production line. This system is currently working in edge environment and its being updated to a cloud native solution. The evaluation shown in Table 15 is done by two individuals from different perspectives. One is from global perspective, and the other is from local perspective. Both evaluations have been done with the new update in mind and whether it should be placed on the Cloud, as it is designed to be implemented in. This decision has been made due to the solution complexity that would come if the track and trace system would be implemented in on-prem solution based on earlier experience with the same type of Kubernetes implementation. In this use-case example, the solution provider has been put out to tender, and the desired solution provider has been selected. Regarding the global evaluation in Table 15, the suggestion of the decision matrix is either one of the solution options based on the fact that both solutions are able to pass the requirements placed on the system and local evaluation gravitates towards edge solution based on the previous experience with the current system and the process requirements associated with the production line and the manufacturing process.

Table 15. User defined importance – Track and Trace.

| **Characteristic** | **USER defined importance (scale 0-4)** | |
|---|---|---|
| | **Global** | **Local** |
| Access Control & Audit Trail | 2 | 4 |
| Availability | 3 | 4 |
| Cost efficiency | 2 | 2 |
| Data Privacy | 3 | 4 |
| GMP Environment | 3 | 0 |
| Infrastructure Complexity | 1 | 0 |
| Latency Sensitivity | 1 | 3 |
| Low Maintenance | 4 | 3 |
| Network Reliability | 2 | 4 |
| Processing Power | 1 | 1 |
| Scalability | 3 | 1 |
| Security | 3 | 4 |
| Total Cost of Ownership | 1 | 2 |
| **Suggestion** | **CLOUD/EDGE** | **EDGE** |

Based on these two evaluations, it can be seen that there is difference on how people view the importance of each characteristic depending on what perspective they are looking from and what drives their decisions. The decision matrix supports both, the already decided cloud implementation for the update coming with the global evaluation, and also the edge solution that the local site values more.

### 4.2.3 Innovation venture

The innovation venture is lager Business Finland funded project that aims to boost digital solutions adaption in life science manufacturing in Finland, and this thesis is part of the target company's development project under the innovation venture umbrella, from which theses use-case come. In this project, two use-cases are implemented and they both fit perfectly for the evaluation of the decision matrix. They both are already existing calculations. Both of these use-cases are presented in Table 16 and the suggested solution is shown individually for both at the end of the table. This table and the importance of each characteristic was evaluated by project manager of the innovation venture who has been working closely to the project where these use-cases have been taken.

Currently distance calculation is implemented in edge and heat exposure is implemented in both cloud and edge. The distance calculation is a conversion of time to distance calculation, and it gives absolute value as a result, and it is not computationally intense. The heat exposure calculation model aims for predictiveness, and it is simple calculation that tries to measure the effect of the heat to the surface quality which is not computational burden in the edge device. The use of the results is the key element that drives the decision regarding the implementation. If the result is needed in real time, edge is the way to go but when there is no real-time requirement for the value, cloud is good solution and in there the value can be used in further processing or such usage. Also, the implementation of individual calculations would be easier to carry out in cloud if the data is already sent there due to change management and validation burden in validated computerized system when the software is being modified and features being added.

Table 16. User defined importance – Innovation venture.

| **Characteristic** | **USER defined importance (scale 0-4)** | |
|---|---|---|
| | **Distance calculation** | **Heat exposure** |
| Access Control & Audit Trail | 2 | 3 |
| Availability | 3 | 1 |
| Cost efficiency | 1 | 1 |
| Data Privacy | 1 | 1 |

| **Characteristic** | **USER defined importance (scale 0-4)** | |
|---|---|---|
| | **Distance calculation** | **Heat exposure** |
| GMP Environment | 3 | 2 |
| Infrastructure Complexity | 1 | 1 |
| Latency Sensitivity | 3 | 3 |
| Low Maintenance | 4 | 3 |
| Network Reliability | 3 | 1 |
| Processing Power | 1 | 1 |
| Scalability | 1 | 1 |
| Security | 1 | 2 |
| Total Cost of Ownership | 2 | 1 |
| **Suggestion** | **CLOUD/EDGE** | **CLOUD/EDGE** |

Based on the results that the decision matrix shows, both of the use-cases would be feasible in either cloud or edge solution. These results align with the view of the project manager and the actual implementation of the heat exposure, and with the distance calculation the implementation would also be possible to be carried out in the cloud which is already existing environment and the data is sent there.

# 5 Conclusions and discussion

This thesis implemented mixed-method research with an aim to answer two research questions regarding  how to evaluate when pharma related data is better to be processed in on-prem edge solution or in cloud solution, and would it be possible and needful to create a standardised guideline to help in the decision-making process.

The aim of the structure of the thesis was designed to guide the reader easily through the regulation aspects the working environment, then to go through the comparable solutions for the computation implementation (cloud and edge) and quickly outlie existing prospects of fog computing and LLMs. After this, move to the survey and interview chapter which plays key role in the development of the next chapter, where the throughput of this thesis is presented and then tested against existing use-cases and finally have the conclusions presented where the implementation and the results are discussed and future development and research ideas are brought to the reader's attention.

The answer to the first research question is obtained from multiple stages in this thesis. The trail to the answer starts from the literature review from which the survey is built upon. Characteristics that arose from the literature review are deductively compiled and the survey questions are designed based on those to answer the first research question (see Chapter 1 and Chapter 3). To get the best result and answer, an interview was conducted with most of the respondents to enrich the outcome of the survey. From the interviews, a critical insight was obtained that was not gained from the survey by itself. This helped to build the throughput of this thesis, the decision-making guideline and the decision matrix (see Appendix 4 on page 102 and Chapter 4), and to inductively weight the individual characteristics in the matrix that are relevant for the decision-making process in a regulated pharma environment. Chapter 4 also answers the second research question. It is possible to build and highly needed to have a standardized decision-making guideline to support in the decision-making process. By having this standardized guideline, businesses benefit such as efficiency and consistency are gained.

The process of building the decision matrix was not as straightforward or easy as thought at the start of the research and writing process. The environment in which this thesis operates – the pharma environment – brings its own constrains and challenges to the table, and depending on the literature source that was used, the way that individual characteristics were classified as being a "benefit" or a "limitation" of the given solution varied based on the perspective and the aim of the research paper. In terms of the final result, larger sample size of the survey focus group, would have given more comprehensive outcome in my opinion. Despite the small sample size, the focus group was precenting variety of professionals and competence to gain non-biased results, and this was successfully achieved, resulting the development of the decision matrix.

The final version of the decision matrix, that was built based on deductive and inductive interpretation of the literature, the survey and the interviews, was tested against ongoing and/or completed projects where some sort of evaluation was already done whether to implement the given application to on-prem solution or to cloud solution. These three use-cases (see Chapters 4.2.1, 4.2.2 & 4.2.3) demonstrate and validate that the decision matrix can be used to determine whether pharma regulated data is better to be processed in cloud over edge or vice versa regardless of the size of the application/project. Each of the case examples are in different scale, MES project is the largest, trac and trace the second largest and innovation ventures examples are the smallest. Results of the evaluation against use-cases shows that the matrix is capable of helping in the decision-making process regardless of the size of the application and the project. It was also noted that the matrix did not always suggest the same solution to be used what the project had chosen and the reasons for this varied. One withdrawal from this evaluation was that the weights of the individual characteristics play key role on the end result of the evaluation and by adjusting those more, the matrix is capable of giving more accurate suggestions. Another key aspect that was noted during the use-case evaluation was that the characteristics could be broken down into more detailed and specific subgroups to prevent the possible misunderstanding of the individual characteristic and its intended meaning. Also, a definitions appendix, similar to one seen in Appendix 3 on page 100 would give further clarification regarding individual characteristic, and this would help the decision maker to understand each characteristic when doing the evaluation. Otherwise, the outcome and the suggestions that the matrix was able to provide were found to be in line with the persons own opinions and expectations regarding what should be the better solution based on the evaluated characteristics.

This thesis includes two additional chapters (2.4 & 2.5) under the literature review that did not contribute to the development of the decision-making guideline but were perceived as giving value to the research and the end result of this work, and that they fill the deficiencies that arose during the research and writing process. The fog computing chapter (2.4) was seen as an addition and supplement to the polarized comparison over cloud versus edge, and that it would be a good topic for future research and development regarding the throughput of this thesis. Fog computing offers benefits such as enabling more accessible computing resources, it addresses infrastructure challenges and constrains that are e.g., bandwidth, connectivity and latency. The LLM chapter (2.5) was considered to broaden the possibilities of having an AI tool to help with the decision-making process with suppose a possible fine-tuned agent for the specific environment that would help to evaluate the best solution of computation for individual applications. This is also something that in the future might be great topic for further and deeper research and study.

## References

Aaltonen, P., & Kurvinen, E. (Eds.). (2025). *Contemporary Issues in Industry 5.0: Towards an AI Integrated Society*. Springer Nature. https://directory.doabooks.org/handle/20.500.12854/152119

*About | FDA*. (2024, June 14). FDA. Retrieved February 7, 2025, from https://www.fda.gov/about-fda

*About us | European Medicines Agency (EMA)*. (2016, August 11). https://www.ema.europa.eu/en/about-us

*About us | Fimea*. (2022). Retrieved February 7, 2025, from https://fimea.fi/en/about_us

*About Us | ISPE*. (n.d.). Retrieved September 4, 2025, from https://ispe.org/about

*Acronyms and Abbreviations | FDA*. (n.d.). Retrieved February 15, 2025, from https://www.accessdata.fda.gov/scripts/cder/acronyms/index.cfm

*Advantages Of Cloud Computing | Google Cloud*. (n.d.). Google Cloud. Retrieved April 25, 2025, from https://cloud.google.com/learn/advantages-of-cloud-computing

*AI Act | Shaping Europe's digital future*. (2025, May 28). Retrieved June 1, 2025, from https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

Al-Dulaimy, A., Sharma, Y., Khan, M. G., & Taheri, J. (2020). *Introduction to edge computing | Edge Computing: Models, technologies and applications*. https://digital-library-theiet-org.ezproxy.utu.fi/doi/abs/10.1049/PBPC033E_ch1

Antonopoulos, N., & Gillam, L. (2017). *Cloud computing* (2nd ed.). Springer.

Banafa, A. (2023). *Introduction to Internet of Things (IoT)*. River Publishers. https://app.knovel.com/hotlink/pdf/id:kt013DYVV2/introduction-internet/internet-things-third

*Baseline Guide Vol 8: Pharma 4.0: Vol. vol 8* (1st edition). (2023). ISPE.

Bhattacharya, S. (2022, April 7). *Edge computing in the pharma manufacturing | HCLTech*. https://www.hcltech.com/blogs/edge-computing-pharma-manufacturing

Braun, A. (2019, January 25). History of IoT: A Timeline of Development. *IoT Tech Trends*. https://www.iottechtrends.com/history-of-iot/

Brozzi, R., Forti, D., Rauch, E., & Matt, D. (2020). The Advantages of Industry 4.0 Applications for Sustainability: Results from a Sample of Manufacturing Companies. *Sustainability*, *12*, 3647. https://doi.org/10.3390/su12093647

Buntz, B. (2021, March 2). *Pharma's been slow to adopt Industry 4.0—But that could change*. Pharmaceutical Processing World. https://www.pharmaceuticalprocessingworld.com/pharmas-been-slow-to-adopt-industry-4-0-but-that-could-change/

Caetano, B. D. L. (2023, July 13). *EU Annex 11: Computerized Systems (What You Need to Know)*. SimplerQMS. https://simplerqms.com/eu-annex-11/

Cao, J., Zhang, Q., & Shi, W. (2018). *Edge Computing: A Primer*. Springer International Publishing. https://doi.org/10.1007/978-3-030-02083-5

Cao, K., Liu, Y., Meng, G., & Sun, Q. (2020). An Overview on Edge Computing Research. *IEEE Access*, *8*, 85714–85728. https://doi.org/10.1109/ACCESS.2020.2991734

Chang, W. L., Grady, N., & Group, N.-P. N. B. D. P. W. (2018). NIST Big Data Interoperability Framework: Volume 1, Big Data Definitions [Version 2]. *NIST*. https://www.nist.gov/publications/nist-big-data-interoperability-framework-volume-1-big-data-definitions-version-2

Chaudhary, S., Somani, G., & Buyya, R. (Eds.). (2017). *Research Advances in Cloud Computing*. Springer. https://doi.org/10.1007/978-981-10-5026-8

Chopra, R. (2017). *Cloud Computing: An Introduction*. Mercury Learning & Information. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=4895092

Choudhari, S., S., Gupta, A., & Kamble, N. (2021). *Security and Privacy of AWS S3*. https://doi.org/10.15680/IJIRSET.2021.1012067|

Choudhary, A., A., Verma, P., & Rai, P. (2021). A walkthrough of Amazon Elastic Compute Cloud (Amazon EC2): A Review. *International Journal for Research in Applied Science and Engineering Technology*, *9*, 93–97. https://doi.org/10.22214/ijraset.2021.38764

Clancy, R. (2023, March 15). What Is Fog Computing? Definition, Applications, Everything to Know. *Cybersecurity Exchange*. https://www.eccouncil.org/cybersecurity-exchange/ethical-hacking/fog-computing-everything-to-know/

*Cloud Computing in Pharma | Viseven*. (2024, September 18). Viseven. Retrieved March 7, 2025, from http://viseven.com/cloud-computing-in-pharma/

*Cognitive computing | Google*. (n.d.). Google Cloud. Retrieved March 14, 2025, from https://cloud.google.com/discover/what-is-cognitive-computing

Creswell, J. W. (with Creswell, J. D.). (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (Fifth edition.). SAGE.

D'Agostino, D., Morganti, L., Corni, E., Cesini, D., & Merelli, I. (2019). Combining Edge and Cloud computing for low-power, cost-effective metagenomics analysis. *Future Generation Computer Systems*, *90*, 79–85. https://doi.org/10.1016/j.future.2018.07.036

Deb, M., & Choudhury, A. (2021). *Hybrid Cloud: A New Paradigm in Cloud Computing* (pp. 1–23). https://doi.org/10.1002/9781119764113.ch1

*Edge Computing Consortium is Established to Deepen Digital Transformation—Huawei press center*. (2016). Huawei. //www.huawei.com/en/news/2016/11/edge-computing-consortium-established

*Edge Computing Reference Architecture 2.0*. (n.d.). Retrieved July 15, 2025, from http://en.ecconsortium.net/Uploads/file/20180328/1522232376480704.pdf

*Eight Emerging Trends Shaping the Future of Cloud Computing | Forbes*. (2024). Forbes. https://www.forbes.com/councils/forbestechcouncil/2024/09/05/eight-emerging-trends-shaping-the-future-of-cloud-computing/

Eigner, E., & Händler, T. (2024). *Determinants of LLM-assisted Decision-Making*. https://doi.org/10.48550/arXiv.2402.17385

Elin, M. N. Z. (2023). Comparative Analysis of Humans and Large Language Models Decision-Making Abilities: Exploring the Potential Considerations Regarding Utilization of Artificial Intelligence in Decision Support Systems. *Journal of Artificial Intelligence & Cloud Computing*, *2*(2), 1–4. https://doi.org/10.47363/JAICC/2023(2)116

*EMA Guideline on computerised systems and electronic data in clinical trials*. (2023).

*Empowering your regulatory and compliance journey | Google Cloud*. (2024). Google Cloud Blog. https://cloud.google.com/blog/products/identity-security/the-power-of-choice-empowering-your-regulatory-and-compliance-journey

*EU AI Act: First regulation on artificial intelligence*. (2023, June 8). Topics | European Parliament. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

EudraLex Volume 4—Annex 11. (2011). In L. Steinborn, *GMP/ISO Quality Audit Manual for Healthcare Manufacturers and Their Suppliers, (Volume 2—Regulations, Standards, and Guidelines)* (0 ed., pp. 257–316). CRC Press. https://doi.org/10.3109/9780203026656-14

*EudraLex—European Commission*. (2025, September 3). https://health.ec.europa.eu/medicinal-products/eudralex_en

*EudraLex—Volume 4—European Commission*. (2025, January 24). https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-4_en

*European Medicines Agency – EMA | European Union*. (2024, October 11). Retrieved January 23, 2025, from https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/search-all-eu-institutions-and-bodies/european-medicines-agency-ema_en

Facts About the Current Good Manufacturing Practice (cGMP) | FDA. (2025). *FDA*. https://www.fda.gov/drugs/pharmaceutical-quality-resources/facts-about-current-good-manufacturing-practice-cgmp

*FAR-EDGE Reference Architecture – Edge4Industry*. (2018, February 12). https://www.edge4industry.eu/knowledge-base/articles/far-edge-reference-architecture/

Ferrer, A., García Pérez, D., & González, R. (2016). Multi-cloud Platform-as-a-service Model, Functionalities and Approaches. *Procedia Computer Science*, *97*, 63–72. https://doi.org/10.1016/j.procs.2016.08.281

*Fog Computing and Internet of Things | Techplayon*. (2017). Retrieved August 5, 2025, from https://www.techplayon.com/fog-computing-and-internet-of-things/

*Fog Computing in the Internet of Things: Intelligence at the Edge | SpringerLink*. (n.d.). Retrieved September 18, 2024, from https://link-springer-com.ezproxy.utu.fi/book/10.1007/978-3-319-57639-8

Foote, K. D. (2021, December 17). A Brief History of Cloud Computing. *DATAVERSITY*. https://www.dataversity.net/brief-history-cloud-computing/

Foster, I., & Kesselman, C. (1999). *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann Publishers.

Furht, B., Escalante, A., & SpringerLink. (2010). *Handbook of Cloud Computing*. Springer US. https://doi.org/10.1007/978-1-4419-6524-0

Gaca, A. (2023, July 11). *How to implement cloud computing | Future Processing*. https://www.future-processing.com/blog/how-to-implement-cloud-computing/

Gajdzik, B., & Wolniak, R. (2022). Influence of Industry 4.0 Projects on Business Operations: Literature and Empirical Pilot Studies Based on Case Studies in Poland. *Journal of Open Innovation: Technology, Market, and Complexity*, *8*(1), 44. https://doi.org/10.3390/joitmc8010044

Gambier, S. (2025, September 1). *Presentation of the throughput of the thesis.* [Personal communication].

*GAMP RDI Good Practice Guide: Data Integrity by Design*. (2020). ISPE.

Gams, M., & Kolenik, T. (2021). Relations between Electronics, Artificial Intelligence and Information Society through Information Society Rules. *Electronics*, *10*, 514. https://doi.org/10.3390/electronics10040514

George, A. S., & George, A. s. (2021). The Evolution of Content Delivery Network: How it Enhances Video Services, Streaming, Games, e-commerce, and Advertising. *International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering*, *10*, 10435–10442. https://doi.org/10.5281/zenodo.6788660

*Global data center spending | Statista*. (2024). Statista. Retrieved March 7, 2025, from https://www.statista.com/statistics/1114926/enterprise-spending-cloud-and-data-centers/

*Global public cloud platforms market | Statista*. (2022). Statista. Retrieved March 7, 2025, from https://www.statista.com/statistics/510340/worldwide-public-cloud-platform-as-a-service/

*Glossary | NIST*. (n.d.). Retrieved December 31, 2024, from https://csrc.nist.gov/glossary?

Golec, M., Hatay, E. S., Golec, M., Uyar, M., Golec, M., & Gill, S. S. (2024). Quantum cloud computing: Trends and challenges. *Journal of Economy and Technology*, *2*, 190–199. https://doi.org/10.1016/j.ject.2024.05.001

*Good manufacturing practice | European Medicines Agency (EMA)*. (2022, April 6). Retrieved February 15, 2025, from https://www.ema.europa.eu/en/human-regulatory-overview/research-development/compliance-research-development/good-manufacturing-practice

Gordan, M., Ghaedi, K., & Saleh, V. (Eds.). (2023). *Industry 4.0: Perspectives and Applications*. IntechOpen. https://directory.doabooks.org/handle/20.500.12854/113219

Gorton, I., Greenfield, P., Szalay, A., & Williams, R. (2008). Data-Intensive Computing in the 21st Century. *Computer*, *41*(4), 30–32. https://doi.org/10.1109/MC.2008.122

Gumilar, K. E., Indraprasta, B. R., Faridzi, A. S., Wibowo, B. M., Herlambang, A., Rahestyningtyas, E., Irawan, B., Tambunan, Z., Bustomi, A. F., Brahmantara, B. N., Yu, Z.-Y., Hsu, Y.-C., Pramuditya, H., Putra, V. G. E., Nugroho, H., Mulawardhana, P., Tjokroprawiro, B. A., Hedianto, T., Ibrahim, I. H., … Tan, M. (2024). Assessment of Large Language Models (LLMs) in decision-making support for gynecologic oncology. *Computational and Structural Biotechnology Journal*, *23*, 4019–4026. https://doi.org/10.1016/j.csbj.2024.10.050

*Health Insurance Portability and Accountability Act of 1996*. (n.d.). Retrieved August 8, 2025, from https://www.govinfo.gov/content/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf

Hoareau, C., & Satoh, I. (2009). Modeling and Processing Information for Context-Aware Computing: A Survey. *New Generation Computing*, *27*(3), 177–196. https://doi.org/10.1007/s00354-009-0060-5

*How integrating edge with cloud, data and AI unlocks value*. (2023). https://www.accenture.com/sg-en/blogs/cloud/edge-adoption

*How Tech Evolved in the 2010's | Global X*. (2020). Global X ETFs. Retrieved March 7, 2025, from https://www.globalxetfs.com/a-decade-of-change-how-tech-evolved-in-the-2010s-and-whats-in-store-for-the-2020s/

Huawei Technologies Co., Ltd., Ltd. (2022). *Cloud Computing Technology*. Springer. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=7105593

Hurwitz, J. S., Kaufman, M., & Bowles, A. (2015). *Cognitive Computing and Big Data Analytics*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=1895129

*IBM Virtual Machine Facility/370*. (1973). https://www.vm.ibm.com/history/50th/vm370ref.pdf

*Industry 5.0—European Commission*. (2024, July 25). https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en

*Investing in Cloud, Edge and the Internet of Things | Shaping Europe's digital future*. (n.d.). Retrieved July 23, 2025, from https://digital-strategy.ec.europa.eu/en/policies/iot-investing

Iorga, M., Feldman, L., Barton, R., Martin, M. J., Goren, N., & Mahmoudi, C. (2018). *Fog computing conceptual model* (No. NIST SP 500-325; p. NIST SP 500-325). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.500-325

*ISPE GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems (Second Edition)* (2nd ed.). (2022). ISBN 978-1-946964-57-1

Jackson, K. L., & Goessling, S. (2018). *Architecting Cloud Computing Solutions: Build Cloud Strategies That Align Technology and Economics While Effectively Managing Risk*. Packt Publishing, Limited. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=5405689

Kaswan, K. S., Dhatterwal, J. S., Jaglan, V., Balusamy, B., & Sood, K. (2024). *Enabling Technologies for Smart Fog Computing*. Institution of Engineering and Technology (The IET). https://app.knovel.com/hotlink/pdf/id:kt0141DVN1/enabling-technologies/history-fog-computing

Kazeem, A. (2018, December 31). *What Is Edge Computing? Examples & Use Cases Explained*. Cloudwards. https://www.cloudwards.net/what-is-edge-computing/

Kiel, D., Müller, J., Arnold, C., & VOIGT, K.-I. (2017). SUSTAINABLE INDUSTRIAL VALUE CREATION: BENEFITS AND CHALLENGES OF INDUSTRY 4.0. *International Journal of Innovation Management*, *21*, 1740015. https://doi.org/10.1142/S1363919617400151

Lin, S.-W., & Simmon, E. (2022). *The Industrial Internet Reference Architecture*.

Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., & Leaf, D. (2011). *NIST Cloud Computing Reference Architecture*.

Luthria, G., & Wang, Q. (2020). Implementing a Cloud Based Method for Protected Clinical Trial Data Sharing. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, *25*, 647–658.

Lynn, T., Mooney, J. G., Rosati, P., & Fox, G. (Eds.). (2020). *Measuring the Business Value of Cloud Computing*. Springer International Publishing. https://doi.org/10.1007/978-3-030-43198-3

Maayan, G. D. (2021, April 6). *How the Cloud Has Evolved Over the Past 10 Years*. DATAVERSITY. https://www.dataversity.net/how-the-cloud-has-evolved-over-the-past-10-years/

Maciel, P., Dantas, J., Melo, C., Pereira, P., Oliveira, F., Araujo, J., & Matos, R. (2022). A survey on reliability and availability modeling of edge, fog, and cloud computing. *Journal of Reliable Intelligent Environments*, *8*(3), 227–245. https://doi.org/10.1007/s40860-021-00154-1

Mannanuddin, K., Kumar, M. R., Aluvala, S., Nagender, Y., & Vishali, S. (2020). Fundamental Perception of EDGE Computing. *IOP Conference Series. Materials Science and Engineering*, *981*(2). https://doi.org/10.1088/1757-899X/981/2/022035

Margulius, D. (2002). Apps on the edge. *InfoWorld*. https://www.infoworld.com/article/2224335/apps-on-the-edge.html

*Marketing authorisations | suomi.fi*. (2025). Retrieved September 4, 2025, from https://www.suomi.fi/services/marketing-authorisations-and-registrations-of-medicinal-products-finnish-medicines-agency-fimea/2335e660-bce4-4622-aae5-2f4ab0af6c99

Matana, T. (2024). Origin & History of ASP SaaS, PaaS and Cloud Computing. *International Journal of Scientific Research and Management (IJSRM)*, *12*, 1571–1607. https://doi.org/10.18535/ijsrm/v12i10.ec08

Meccyjoy, Mohammed, Mark, M., Soetan, T., & Ahmed, M. (2024). Unveiling the Future: How Large Language Models Are Shaping Smart Decision-Making. *Artificial Intelligence Review*, *7*, 12.

Mei, L. (2022). *Fintech Fundamentals: Big Data / Cloud Computing / Digital Economy*. Mercury Learning & Information. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=7021248

Mell, P., & Grance, T. (2011). *The NIST Definition of Cloud Computing*.

*Merging Cloud with Edge Computing | LinkedIn*. (2024). Retrieved March 7, 2025, from https://www.linkedin.com/pulse/merging-cloud-edge-computing-future-data-processing-mubashir-aijaz-g4ezf/

Mohammed, S., Fang, W. C., & Ramos, C. (2023). Special issue on '"artificial intelligence in cloud computing."' *Computing*, *105*(3), 507–511. https://doi.org/10.1007/s00607-021-00985-z

Najana, M., & Ranjan, P. (2024). *Compliance and Regulatory Challenges in Cloud Computing: A Sector-Wise Analysis*. *3*. https://doi.org/10.21428/e90189c8.68b5dea5

Nousala, S., Metcalf, G., & Ing, D. (2024). *Industry 4. 0 to Industry 5. 0: Explorations in the Transition from a Techno-Economic to a Socio-technical Future*. Springer Singapore Pte. Limited. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=31281953

*OpenFog | OPCfoundation.org*. (n.d.). Retrieved September 11, 2025, from https://opcfoundation.org/markets-collaboration/openfog/

Packard, N. (2023). INTERNET Prehistory: ARPANET Chronology. *Cogent Social Sciences*, *9*(2), 2245237. https://doi.org/10.1080/23311886.2023.2245237

Papakostas, N., Constantinescu, C., & Mourtzis, D. (Eds.). (2020). *Novel Industry 4.0 Technologies and Applications*. MDPI - Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/books978-3-03943-584-5

Patel, M., Sabella, D., Sprecher, N., & Young, V. (2015). *Mobile edge computing—A key technology towards 5G*.

Pérez, J., Díaz, J., Berrocal, J., López-Viana, R., & González-Prieto, Á. (2022). Edge computing. *Computing*, *104*(12), 2711–2747. https://doi.org/10.1007/s00607-022-01104-2

*Pharma 4.0 for plant manufacturing: A roadmap to the future | Schneider Electric*. (n.d.). Retrieved March 21, 2025, from https://www.se.com/fi/fi/download/document/Pharma4.0_eGu_/

*Pharma 4.0 Operating Model | Industry 4.0 | ISPE | International Society for Pharmaceutical Engineering*. (n.d.). Retrieved September 6, 2025, from https://ispe.org/initiatives/pharma-4.0

*Pharma 4.0TM | ISPE*. (2024, February 14). https://ispe.org/topics/pharma-4.0

*Pharmaceutical industry | Definition, Overview, History, Characteristics, Examples, & Facts | Britannica*. (2025, January 14). Retrieved January 17, 2025, from https://www.britannica.com/technology/pharmaceutical-industry

Pharmaceutical Inspections and Compliance | FDA. (2024). *FDA*. Retrieved February 7, 2025, from https://www.fda.gov/drugs/guidance-compliance-regulatory-information/pharmaceutical-inspections-and-compliance

*Pharmacovigilance | Fimea*. (n.d.). Fimea. Retrieved September 4, 2025, from https://fimea.fi/en/supervision/pharmacovigilance

*PIC/S Secretariat—PI 011-3*. (2007, September).

*Public cloud computing market size | Statista*. (2025). Statista. Retrieved September 9, 2025, from https://www.statista.com/statistics/273818/global-revenue-generated-with-cloud-computing-since-2009/

Pushpa, J., & Kalyani, S. A. (2020). Chapter Three—Using fog computing/edge computing to leverage Digital Twin. In P. Raj & P. Evangeline (Eds.), *Advances in Computers* (Vol. 117, pp. 51–77). Elsevier. https://doi.org/10.1016/bs.adcom.2019.09.003

Rahmani, A. M., Liljeberg, P., Preden, J.-S., & Jantsch, A. (Eds.). (2018). *Fog Computing in the Internet of Things*. Springer International Publishing. https://doi.org/10.1007/978-3-319-57639-8

*RAMI 4.0 | isa.org*. (n.d.). Isa.Org. Retrieved March 13, 2025, from https://www.isa.org/intech-home/2019/march-april/features/rami-4-0-reference-architectural-model-for-industr

Reddy, N. (2019). *A Survey on Specialised Hardware for Machine Learning*. https://doi.org/10.13140/RG.2.2.20697.26725

*Reference Architectural Model Industrie 4.0 | Plattform Industrie 4.0*. (2018). Retrieved March 30, 2025, from https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/rami40-an-introduction.html

*Reflecting on our ten-year App Engine journey*. (2018). Google Cloud Blog. https://cloud.google.com/blog/products/gcp/reflecting-on-our-ten-year-app-engine-journey

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA Relevance) (2024). http://data.europa.eu/eli/reg/2024/1689/oj/eng

*Regulation—2016/679—EN - GDPR - EUR-Lex*. (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

*Reminiscences on the Theory of Time-Sharing*. (n.d.). Retrieved February 27, 2025, from http://jmc.stanford.edu/computing-science/timesharing.html

Rosati, P., & Lynn, T. (2020). Measuring the Business Value of Infrastructure Migration to the Cloud. In T. Lynn, J. G. Mooney, P. Rosati, & G. Fox (Eds.), *Measuring the Business Value of Cloud Computing* (pp. 19–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-43198-3_2

Rountree, D., & Castrillo, I. (2013). *The Basics of Cloud Computing: Understanding the Fundamentals of Cloud Computing in Theory and Practice*. Elsevier Science & Technology Books. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=1115176

Sandoval, K. (2017, May 25). *What is Fog Computing? | Nordic APIs*. Nordic APIs. https://nordicapis.com/what-is-fog-computing/

Sarafe, P. (2023). *The WWW Journey of Web 1.0 to 3.0* (p. 202 P.).

Shapla, R., Rahim, M., Roy, S., Uzzal, K., & Prodhan, Dr. U. (2020). *Cloud Computing: Architecture, Services, Deployment Models, Storage, Benefits and Challenges*. *4*, 837–842.

Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge Computing: Vision and Challenges. *IEEE Internet of Things Journal*, *3*(5), 637–646. https://doi.org/10.1109/JIOT.2016.2579198

Shi, W., Pallis, G., & Xu, Z. (2019). Edge Computing [Scanning the Issue]. *Proceedings of the IEEE*, *107*(8), 1474–1481. Proceedings of the IEEE. https://doi.org/10.1109/JPROC.2019.2928287

Singh, R., & Gill, S. S. (2023). Edge AI: A survey. *Internet of Things and Cyber-Physical Systems*, *3*, 71–92. https://doi.org/10.1016/j.iotcps.2023.02.004

Singh, S. (2017). Optimize cloud computations using edge computing. *2017 International Conference on Big Data, IoT and Data Science (BID)*, 49–53. https://doi.org/10.1109/BID.2017.8336572

Soinio, H. (2025, September 25). *Interview after survey* [Personal communication].

Sony, M., Antony, J., Mc Dermott, O., & Garza-Reyes, J. A. (2021). An empirical examination of benefits, challenges, and critical success factors of industry 4.0 in manufacturing and service sector. *Technology in Society*, *67*, 101754. https://doi.org/10.1016/j.techsoc.2021.101754

Spjuth, O., Frid, J., & Hellander, A. (2021). The machine learning life cycle and the cloud: Implications for drug discovery. *Expert Opinion on Drug Discovery*, *16*(9), 1071–1079. https://doi.org/10.1080/17460441.2021.1932812

*System Requirements for Deep Learning | GeegsforGeeks*. (2024). GeeksforGeeks. Retrieved February 28, 2025, from https://www.geeksforgeeks.org/system-requirements-for-deep-learning/

Taheri, J., & Deng, S. (2020). *Edge computing: Models, technologies and applications*. The Institution of Engineering and Technology.

Tallon, P. P., Mooney, J. G., & Duddek, M. (2020). Measuring the Business Value of IT. In T. Lynn, J. G. Mooney, P. Rosati, & G. Fox (Eds.), *Measuring the Business Value of Cloud Computing* (pp. 1–17). Springer International Publishing. https://doi.org/10.1007/978-3-030-43198-3_1

Taneja, M., & Davy, A. (2016). Resource Aware Placement of Data Analytics Platform in Fog Computing. *Procedia Computer Science*, *97*, 153–156. https://doi.org/10.1016/j.procs.2016.08.295

Thames, L., & Schaefer, D. (2017). *Industry 4.0: An Overview of Key Benefits, Technologies, and Challenges*. https://doi.org/10.1007/978-3-319-50660-9_1

*The evolution of containers | TechTarget*. (2023). Search IT Operations. Retrieved March 7, 2025, from https://www.techtarget.com/searchitoperations/feature/Dive-into-the-decades-long-history-of-container-technology

*The Future of Cloud 2028 | Gartner*. (n.d.). Gartner. Retrieved March 7, 2025, from https://www.gartner.com/en/webinar/566448/1276680

*The Future of Cloud Computing 2025-2030 | CloudDefense.AI*. (2024). Retrieved March 7, 2025, from https://www.clouddefense.ai/future-of-cloud-computing/

*The Future of Cloud Cybersecurity | CSA*. (2024). Retrieved March 7, 2025, from https://cloudsecurityalliance.org/blog/2024/04/29/the-future-of-cloud-cybersecurity

*The history of cloud computing | Cyber magazine*. (2021). Retrieved March 7, 2025, from https://cybermagazine.com/cloud-security/history-cloud-computing

*The history of virtualization and its mark on data center management | TechTarget*. (2019). Search IT Operations. Retrieved March 7, 2025, from https://www.techtarget.com/searchitoperations/feature/The-history-of-virtualization-and-its-mark-on-data-center-management

*The next generation Internet of Things | European Commission*. (n.d.). Retrieved March 7, 2025, from https://digital-strategy.ec.europa.eu/en/policies/next-generation-internet-things

*The Rise of Edge Computing: Transforming the Future of Data Processing—Atlanta Technology Professionals*. (2024, July 1). https://atpconnect.org/the-rise-of-edge-computing-transforming-the-future-of-data-processing/

*Top 10 Cloud Computing Trends | GeeksforGeeks*. (2024). GeeksforGeeks. Retrieved March 7, 2025, from https://www.geeksforgeeks.org/cloud-computing-trends/

Top Cloud Computing News of The Decade | VMWare. (2021, January 21). *VMware Cloud Management*. Retrieved March 7, 2025, from https://blogs.vmware.com/management/2021/01/top-cloud-news-of-the-decade.html

Tu, X., He, Z., Huang, Y., Zhang, Z.-H., Yang, M., & Zhao, J. (2024). An overview of large AI models and their applications. *Visual Intelligence*, *2*(1), 34. https://doi.org/10.1007/s44267-024-00065-8

*United States | European Medicines Agency (EMA)*. (2023). Retrieved January 10, 2025, from https://www.ema.europa.eu/en/partners-networks/international-activities/bilateral-interactions-non-eu-regulators/united-states

*Using FinOps to optimize AI and maximize ROI | DataScienceCentral*. (2025). Data Science Central. Retrieved March 7, 2025, from https://www.datasciencecentral.com/using-finops-to-optimize-ai-and-maximize-roi/

Vahdat Nejad, H., Eilaki, S., & Izadpanah, S. (2018). Towards a Better Understanding of Ubiquitous Cloud Computing. *International Journal of Cloud Applications and Computing*, *8*, 1–20. https://doi.org/10.4018/IJCAC.2018010101

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Wang, H., Zhang, F., & Mu, C. (2025). *One for All: A General Framework of LLMs-based Multi-Criteria Decision Making on Human Expert Level*. https://doi.org/10.48550/arXiv.2502.15778

*What are Large Language Models? | NVIDIA Glossary*. (n.d.). NVIDIA. Retrieved June 1, 2025, from https://www.nvidia.com/en-us/glossary/large-language-models/

*What Are Large Language Models (LLMs)? | IBM*. (2023, November 2). Retrieved June 1, 2025, from https://www.ibm.com/think/topics/large-language-models

*What is AIaaS | Microsoft Azure*. (n.d.). Retrieved February 28, 2025, from https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-aiaas

*What is Cloud Computing Infrastructure | VMWare*. (2024). Retrieved December 31, 2024, from https://www.vmware.com/topics/cloud-computing-infrastructure

What Is Edge Computing? Components, Examples, and Best Practices - Spiceworks. (2025). *Spiceworks Inc*. https://www.spiceworks.com/tech/edge-computing/articles/what-is-edge-computing/

*What is GAMP® | ISPE*. (n.d.). Retrieved February 15, 2025, from https://ispe.org/initiatives/regulatory/what-gamp

*What is GMP? - ECA Academy*. (n.d.). Retrieved February 15, 2025, from https://www.gmp-compliance.org/what-is-gmp

*What is LLM? | AWS*. (n.d.). Amazon Web Services, Inc. Retrieved June 1, 2025, from https://aws.amazon.com/what-is/large-language-model/

*What is reference architecture? | hpe.com*. (n.d.). Retrieved August 6, 2025, from https://www.hpe.com/emea_europe/en/what-is/reference-architecture.html

*What is XaaS | IBM*. (n.d.). Retrieved February 28, 2025, from https://www.ibm.com/think/topics/xaas

*What You Need to Know About GAMP® 5 Guide, 2nd Edition | Pharmaceutical Engineering*. (2025, February 13). Retrieved February 15, 2025, from https://ispe.org/pharmaceutical-engineering/january-february-2023/what-you-need-know-about-gampr-5-guide-2nd-edition

*Who Coined "Cloud Computing"?* (2011). MIT Technology Review. https://www.technologyreview.com/2011/10/31/257406/who-coined-cloud-computing/

Yang, Y., Huang, J., Zhang, T., & Weinman, J. (2020). *Fog and Fogonomics: Challenges and Practices of Fog Computing, Communication, Networking, Strategy, and Economics*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/kutu/detail.action?docID=6020723

Yeung, T. (2019, October 22). What Is Edge Computing? *NVIDIA Blog*. https://blogs.nvidia.com/blog/what-is-edge-computing/

Yi, S., Hao, Z., Qin, Z., & Li, Q. (2015). Fog Computing: Platform and Applications. *2015 Third IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb)*, 73–78. https://doi.org/10.1109/HotWeb.2015.22

## Appendices

**Appendix 1 Trend in 2010’s for cloud market ...................................................... 94**

**Appendix 2 Survey questions and answers – Forms .......................................... 96**

**Appendix 3 Definitions and meanings for survey ............................................. 100**

**Appendix 4 Decision guideline.......................................................................... 102**

## Appendix 1 Trend in 2010’s and 2020’s for cloud market

These graphs demonstrate the trend in enterprise spending on cloud infra and public cloud services in 2010's and 2020’s.

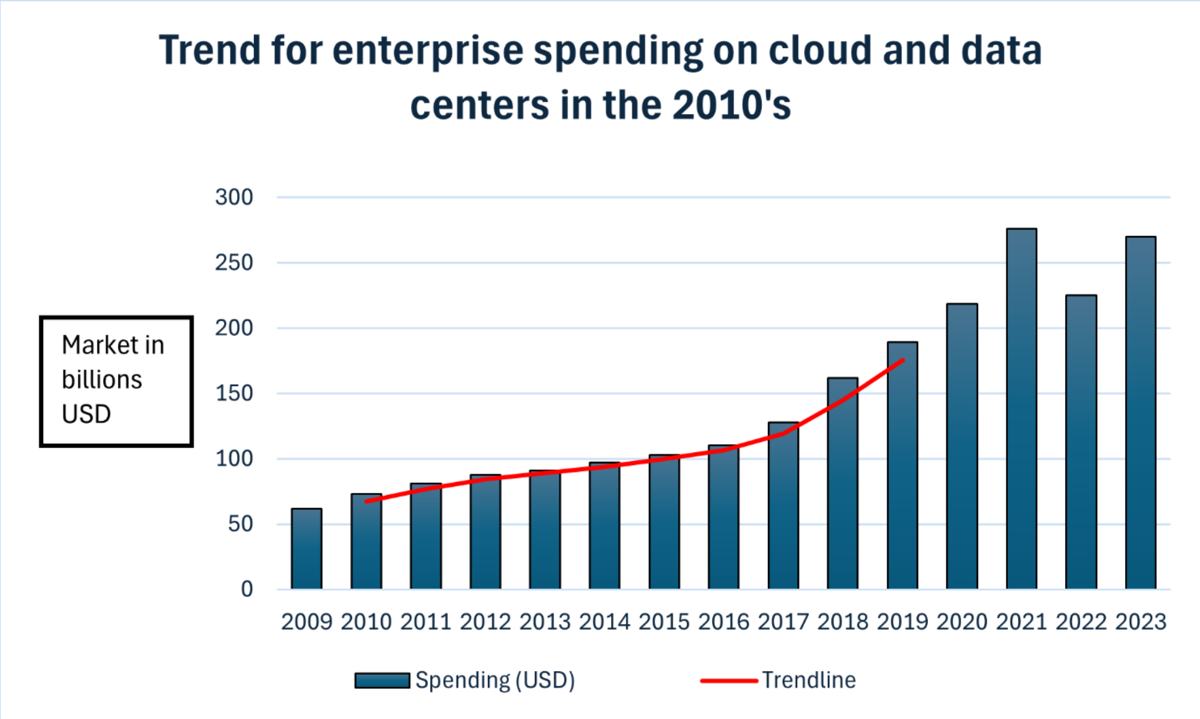



Figure 20. Enterprise spending on cloud datacentres including HW, SW and cloud infrastructure services (*Global Data Centre Spending | Statista, 2024*).

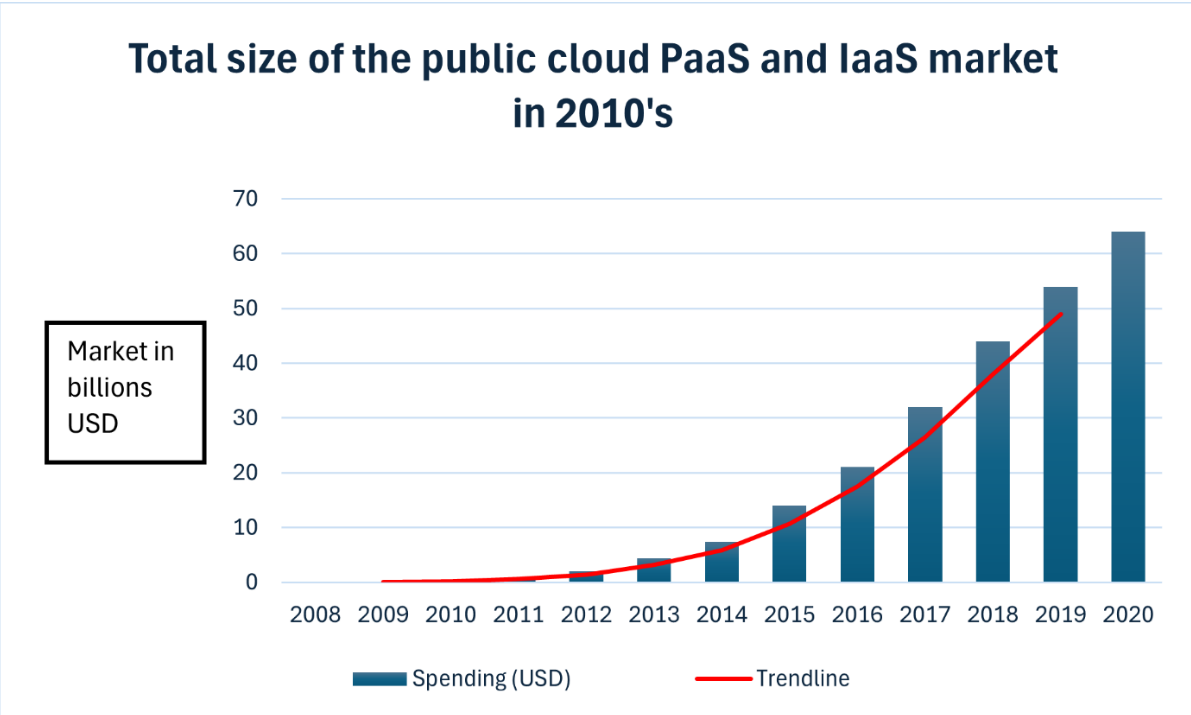



Figure 21. Total size of public cloud market consisting of PaaS and IaaS (*Global Public Cloud Platforms Market | Statista, 2022*).

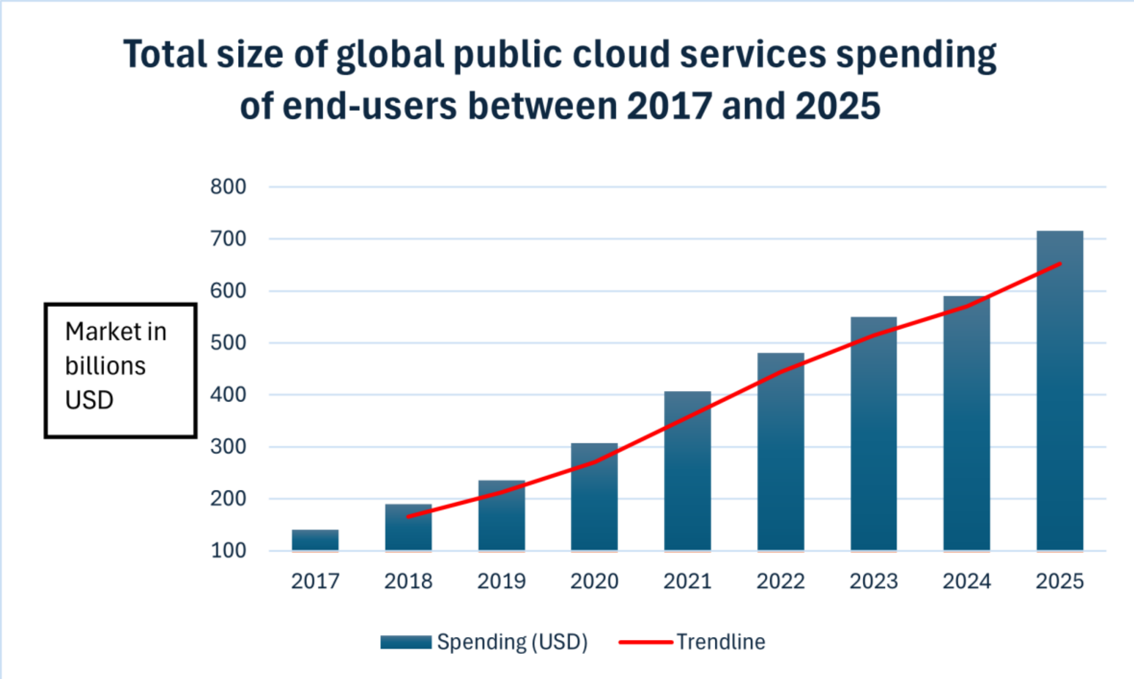
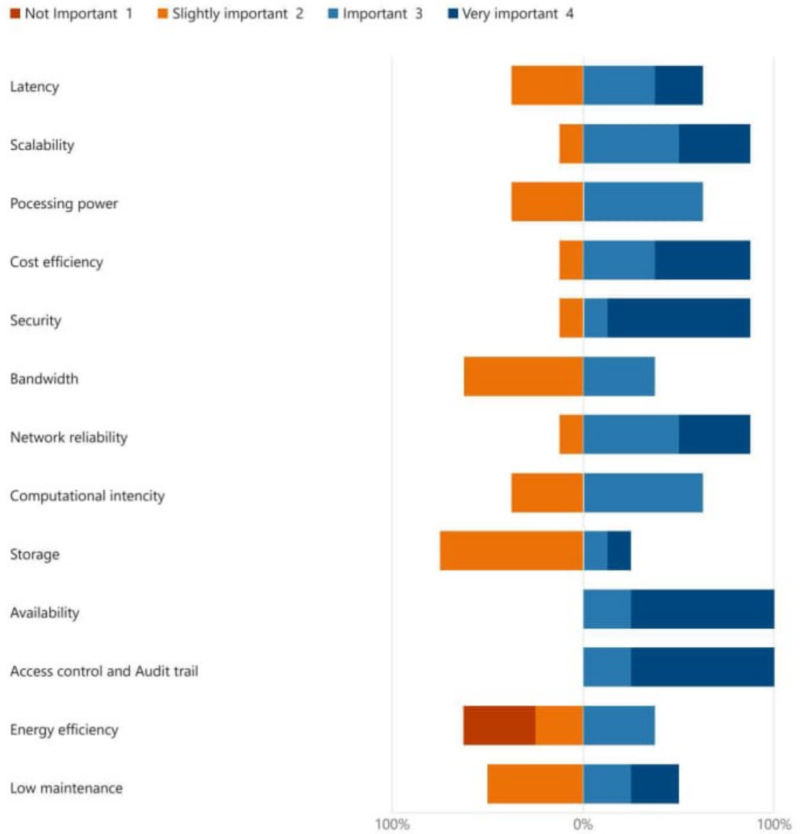


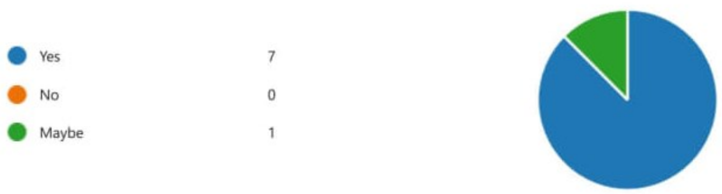

Figure 22. Total size of global spending in public cloud services (*Public Cloud Computing Market Size | Statista, 2025*).

# Appendix 2 Survey questions and answers – Forms

1. How would you rate the importance of these individual characteristics for the decision-making?



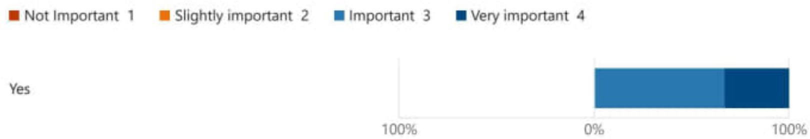

2. Does the pharma environment (GMP) affect the decision regarding data processing location?




3. If YES, please rate the importancy.

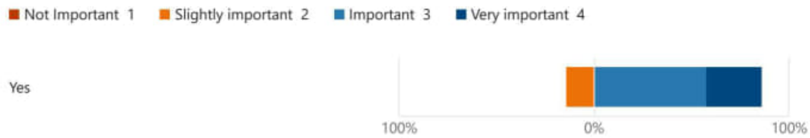
4. Should Digital Twins be considered as a norm in future and would that be important regarding the decision over Cloud/Edge?


5. If YES, please rate the importancy.



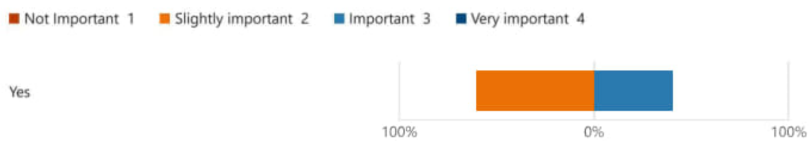

6. Is infrastructure complexity a factor for the decision?


7. If YES, please rate the importancy.




8. Would vendor-lock be an issue?


9. If YES, please rate the importancy.

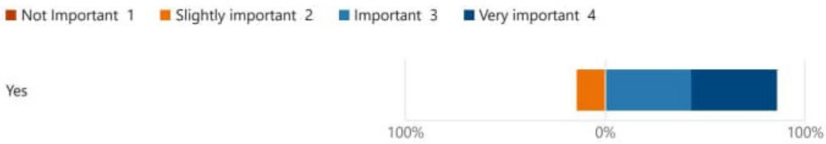


10. Is CapEx vs. OpEx important comparison in the decision?

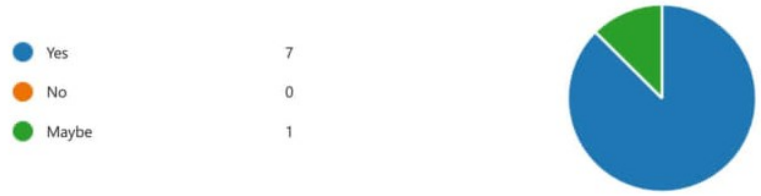

11. If YES, please rate the importancy.


12. Should Total Cost of Ownership (TCO) be considered when doing the decision?


13. If YES, please rate the importancy.

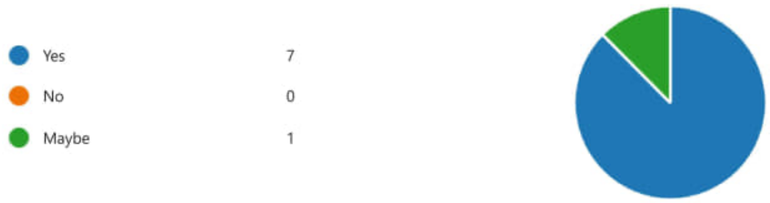



14. Is the ability to synchronize data (between sites and/or cloud services) a relevant consideration?

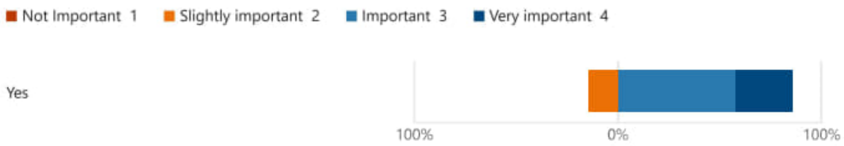



15. If YES, please rate the importancy.

16. Do legacy systems rise concern in integration-wise and should it be an evaluation criteria in the decision ?




17. If YES, please rate the importancy.



# Appendix 3 Definitions and meanings for survey

**Latency**

Network latency is the delay in network communication. It shows the time that data takes to transfer across the network.

→ Lower latency = faster response.

Example: A robot arm on a factory line needs to respond in milliseconds — delay can cause defects.

**Scalability**

Scalability refers to the ability to increase or decrease IT resources as needed to meet changing demand.

→ A scalable system = cost savings and maintaining performance in varying demand

Example: An app that handles 100 users today, and 100,000 tomorrow, needs to scale up fast to remain operational.

**Processing Power**

Processing power refers to the capability of a computer's processor to execute instructions and perform calculations efficiently.

→ More power = faster results for complex tasks.

Example: Analyzing large medical images or running 3D simulations requires strong processing.

**Cost Efficiency**

Cost efficiency means that you effectively manage and optimize costs that are associated with the computing implementation.

→ It's about balancing performance with how much it costs to run.

Example: Edge devices cost more upfront but may save money over time by reducing cloud traffic.

**Security**

Network security is the protection of the underlying networking infrastructure from unauthorized access, misuse, or theft. It involves creating a secure infrastructure for devices, applications, users, and applications to work in a secure manner.

→ Strong security keeps data safe and prevents unauthorized access.

Example: A smart home system encrypts all communications so hackers can't control your locks or see your cameras.

**Bandwidth**

Network bandwidth is a measurement indicating the maximum capacity of a wired or wireless communications link to transmit data over a network connection in a given amount of time.

→ Higher bandwidth = faster, smoother data transfers.

Example: Streaming high-quality video from a drone to the cloud requires lots of bandwidth.

**Network Reliability**

Network reliability refers to the ability of a computer network system to consistently and predictably perform its functions without experiencing failures or interruptions.

→ A reliable network means fewer interruptions, ensuring your system stays connected and functional without unexpected downtime.

Example: In a smart factory, edge devices monitor and control machinery locally so that production doesn't stop even if the connection to the central cloud system is temporarily lost.

**Computational Intensity**

Compute-intensive describes a computer program or process that requires multiple computations to function as designed. Compute-intensive programs require large amounts of computing processing power and memory for proper execution

→ More intense = Need for processing power and specialized hardware like GPUs or TPUs to run efficiently.

Example: In manufacturing, running real-time defect detection using AI on production line images requires powerful processors to analyze data quickly and accurately.

**Storage**

Computer storage refers to components that provide long-term data retention, such as hard disk drives (HDDs) and solid-state drives (SSDs). Storage keeps your data, files, programs, and operating system intact even when the computer is turned off

→ Local (Edge) storage is limited; cloud storage is scalable.

Example: Video surveillance can store short clips locally and send full footage to the cloud.

Sources
https://aws.amazon.com/what-is/latency/
https://www.vmware.com/topics/cloud-scalability
https://www.sciencedirect.com/topics/computer-science/processing-power
https://learn.microsoft.com/en-us/azure/cloud-adoption-framework/strategy/inform/cost-efficiency
https://www.cisco.com/c/en/us/products/security/what-is-network-security.html
https://www.techtarget.com/searchnetworking/definition/bandwidth
https://obkio.com/blog/how-to-measure-improve-network-reliability/
https://www.capterra.com/glossary/compute-intensive/
https://www.hp.com/us-en/shop/tech-takes/computer-memory-vs-storage
https://www.cisco.com/c/en/us/solutions/hybrid-work/what-is-high-availability.html
https://energy-efficient-products.ec.europa.eu/faqs-0/what-energy-efficiency_en
https://www.americaneagle.com/insights/blog/post/what-is-platform-as-a-service-paas-a-comprehensive-guide

**Availability**

High availability means that an IT system, component, or application can operate at a high level, continuously, without intervention, for a given time period. High-availability infrastructure is configured to deliver quality performance and handle different loads and failures with minimal or zero downtime.

→ High availability means it works all the time, with minimal downtime.

Example: An online banking app must be available 24/7 to customers worldwide.

**Access Control and Audit Trail**

Managing who can access the system and keeping a record of their actions.

→ Access control ensures that only authorized users can view or modify data. An audit trail tracks who did what and when, helping with accountability, security, and compliance.

Example: In a manufacturing plant, only certified engineers can adjust machine settings, and every change is logged so managers can review who made which updates and when.

**Energy Efficiency**

The practice of using less energy to provide the same amount of useful output from a.

→ Efficient systems save power and are better for remote or battery-powered environments.

Example: A smart sensor in the forest must run on solar or battery power and can't afford to waste energy.

**Low-Maintenance**

Consisting of infrastructure updates, patching, and maintenance of tools.

→ Low maintenance = less time, lower cost, and lower effort to manage the system.

Example: With low-maintenance solution, it reduces the burden on IT teams, freeing them to other work.

**Infrastructure complexity**

How complicated the system is to set up, manage, and maintain.

→ A more complex infrastructure often needs more planning, skilled staff, and time to manage.

Example: A smart factory using both cloud and edge computing must coordinate servers, networks, security, and software updates across many locations.

**Vendor-lock**

How hard it is to switch to another provider once you've chosen one.

→ Some platforms make it difficult or expensive to move your systems or data elsewhere.

Example: A company using a cloud provider's unique AI tools may find it costly and time-consuming to migrate to a different provider later.

**CapEx (Capital Expenditure)**

Money spent upfront to buy equipment, hardware, or build infrastructure.

→ This is a one-time investment, often higher at the beginning.

Example: Buying edge devices and setting up local servers in a factory is considered CapEx.

**OpEx (Operational Expenditure)**

Ongoing costs to keep the system running, such as subscriptions, maintenance, or electricity.

→ This is paid over time and usually depends on usage.

Example: Monthly payments to a cloud provider for data processing and storage are OpEx.

**Total Cost of Ownership**

The full cost of a system over its entire life — including CapEx + OpEx + maintenance + training.

→ Helps compare long-term value between solutions.

Example: While cloud may have low upfront costs over 5 years, the total cost (TCO) might be higher than an edge solution due to ongoing fees.

https://www.igi-global.com/dictionary/network-complexity/20107
https://www.cloudflare.com/learning/cloud/what-is-vendor-lock-in/
https://www.investopedia.com/terms/c/capitalexpenditure.asp
https://www.investopedia.com/terms/o/operating_expense.asp
https://www.investopedia.com/terms/t/totalcostofownership.asp

# Appendix 4 Decision-making guideline

| | | |
|---|---|---|
| | Decision-making guideline for Cloud vs. Edge computing | Doc.No.: |
| Entity: Universal | Entity No: | Content Type: Decision Guideline |

**Table of Contents**

1. Change history ........ 1
2. Instructions for use ........ 1
3. Decision matrix ........ 2

## 1. Change history

| Version # | Description | Author | Date |
|---|---|---|---|
| **1.0** | Original | Antton Laiho | 14.08.2025 |
| | | | |

## 2. Instructions for use

The decision-making matrix is tool that is designed to help in the decision-making process when comparison between Cloud and Edge solutions are evaluated against each other. Each characteristic is evaluated, and the user-defined importance value is set. The scale is from zero (0) to four (4). Zero is ignored if the person sees that individual characteristics are not relevant to the current application being evaluated, this value is not counted to the total sum. One is “not important” and four is “very important”.

When every characteristic is evaluated, the matrix suggests a computational implementation to go with. This suggestion is either Cloud, Edge or Cloud/Edge, depending on the calculated weighted scores. If the values are within a pre-defined tolerance, then the suggestion is Cloud/Edge, otherwise the solution with the higher score is displayed in the suggestion cell.

Regardless of the solution that the matrix is proposing, the person or the group of people that are making the evaluation have the final word whether to go with cloud or edge solution. This guideline and decision matrix is designed to be a supporting tool for the decision-making that gives suggestion based on inductively weighted values of each characteristic, and not an imperative implementation method tool.

| | | |
|---|---|---|
| | Decision-making guideline for Cloud vs. Edge computing | Doc.No.: |
| Entity: Universal | Entity No: | Content Type: Decision Guideline |

## 3. Decision matrix

*Table 1. Decision matrix*

*User-defined importance scale from 0 to 4.*
*Zero (0) is not taken into account, one (1) being NOT IMPORTANT and four (4) being VERY IMPORTANT.*

| Characteristic | Weigh of the characteristic (%) | USER-defined importance (scale 0-4) | Cloud weighted score | Edge weighted score |
|---|---|---|---|---|
| Access Control & Audit Trail | 9,00 % | | 0 | 0 |
| Availability | 9,00 % | | 0 | 0 |
| Cost efficiency | 8,00 % | | 0 | 0 |
| Data Privacy | 7,00 % | | 0 | 0 |
| GMP Environment | 7,50 % | | 0 | 0 |
| Infrastructure Complexity | 7,00 % | | 0 | 0 |
| Latency Sensitivity | 4,00 % | | 0 | 0 |
| Low Maintenance | 5,00 % | | 0 | 0 |
| Network Reliability | 7,00 % | | 0 | 0 |
| Processing Power | 3,50 % | | 0 | 0 |
| Scalability | 8,50 % | | 0 | 0 |
| Security | 9,50 % | | 0 | 0 |
| Total Cost of Ownership | 15,00 % | | 0 | 0 |
| **Total** | **100 %** | | **0,00** | **0,00** |

Fixed values/ Calculated values

User defined values

Suggested computational implementation to go with

**CLOUD/EDGE**