Seminar Report

Emerging Trends in Cloud Storage

Johann Eilts

MatrNr: 11780368

Supervisor: Chirag Mandal

Georg-August-Universität Göttingen
Institute of Computer Science

March 31, 2024

Abstract

This report explores the evolving subject of cloud storage by presenting three different trends: edge computing, sustainability initiatives, and the application of artificial intelligence in data management. These topics are pivotal in cloud technologies due to the increasing amount of data produced and consumed each year, as well as the need to the reduce the environmental impact of cloud infrastructure and the urgency to improve efficiency when handling the growing amounts of data. A literature research has been conducted to gain insights into the presented areas and provide use cases that show the applicability and relevance of each subject. The report concludes by showing how intertwined these trends are with each other and the importance of the increasing data creation and consumption by organisations and individuals.

Declaration on the use of ChatGPT and comparable tools in the context of examinations

In this work I have used ChatGPT or another AI as follows:

- ☐ Not at all
- ☑ During brainstorming
- ☐ When creating the outline
- ☐ To write individual passages, altogether to the extent of 0% of the entire text
- ☐ For the development of software source texts
- ☐ For optimizing or restructuring software source texts
- ☐ For proofreading or optimizing
- ☑ Further, namely: Literature search assistance

I hereby declare that I have stated all uses completely.
Missing or incorrect information will be considered as an attempt to cheat.

Contents

List of Tables

List of Figures

List of Abbreviations

AI Artificial Intelligence

APIs Application Programming Interfaces

AWS Amazon Web Services

ECC Edge Computing Consortium

ESD Energy Storage Devices

IoT Internet of Things

NIST National Institute of Standards and Technology

NVMe Non-Volatile Memory Express

SATA Serial AT Attachment

1 Introduction

Cloud technologies have changed businesses across diverse industries by offering on-demand access to storage and computing resources, unlimited scaling and flexible pricing for organization and will undoubtedly continue to drive changes in the future [KH20]. Initially the goal of cloud computing was cost reduction through hardware reductions, but nowadays the cloud lets organizations transform their business with various cloud strategies. Automation, standardization and enablement through self-service mechanisms are just the tip of opportunities that cloud technologies offer today [KH20; Lis21]. Lisdorf also compares the transition from on-premise data centers to cloud comparing it to the transition of horse-carriages to automobiles. They also mention that this shift lets companies reduce costs by transitioning from capital expenditure (in-advance, oftentimes difficult to predict investments) to operational expenditure (ongoing, easy to scale based on demand)[Lis21]. Cloud computing therefore enables companies to focus on their core value creation processes, as the planning for storage and computing resource provisioning is reduced to a fraction in comparison to on-premise solutions. [ZLB10]

These are just some of the many advantages that cloud computing and storage offer. To shed light onto ongoing developments in cloud storage, this report will present three emerging trends: (1) Edge Computing that pushes computing tasks from the network core (e.g., data centers) to the network edge [IEE], (2) Sustainability in cloud to reduce the environmental impact of cloud storage [Bal+11] and (3) Artificial Intelligence (AI) in data management [tab]. A literature research in scientific databases (e.g. IEEE Xplore or arxiv.org), as well as on reliable internet sources has been conducted to derive the insights in the following sections. The report will first give a general definition of cloud and cloud storage and then advances to describing each trend and presenting scientific and real-world use cases.

1.1 Cloud Storage Relevance

One of the most important characteristics of cloud storage for organizations and individuals is reducing the burden of building local hardware storage and storage management systems [MGN]. Cloud storage in general presents a major aspect of the whole cloud computing ecosystem, where cloud resources can be conveniently tailored to the use cases of organizations and individuals [KH20; Lis21]. Cloud providers also deploy systems that monitor regulatory aspects of data storage which shifts the responsibility away from organizations [Lis21]. Third parties can also be employed to audit storage systems to ensure adherence to regulations and therefore providing a peace-of-mind to organizations [Wan+13]. Another important aspect are backup-mechanisms accommodated by cloud-storage providers. These also help organizations reduce the stress of full data availability at all times, even in the event of failure [TOM16]. As of 2023, Statista predicts a annual growth rate of 28 % in global data created, captured, copied and consumed. This translates into about 147 Zettabytes in 2024 [Sta23b].

2 Cloud Definitions

As defined by the National Institute of Standards and Technology (NIST) part of the U.S. Department of Commerce, cloud computing is:

> "[...] A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models." [MGN]

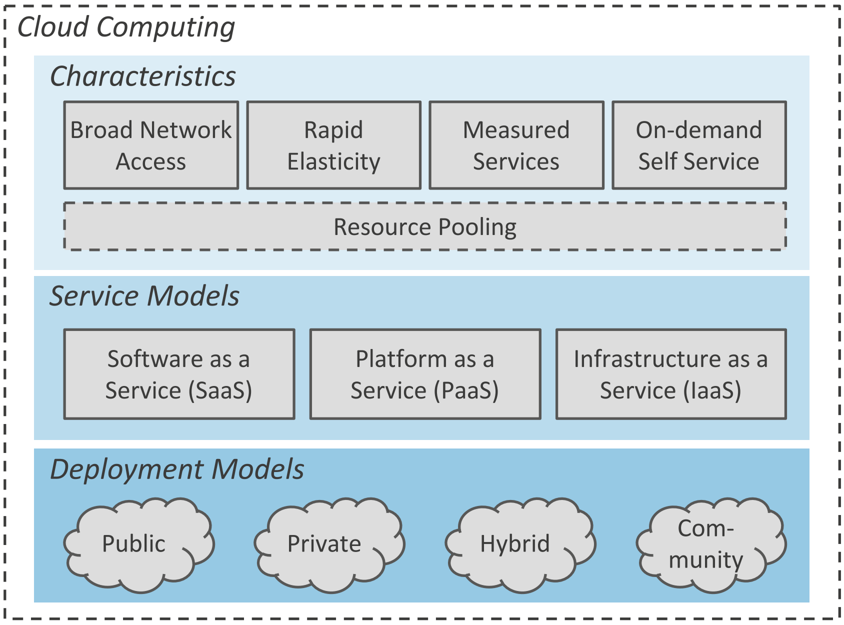

Figure 1: Cloud Computing Definition, own illustration based on Mell et al. [MGN].

Cloud Storage specifically refers to remotely storing data that is accessible via the internet. This also corresponds to the general nature of cloud technology that data can be accessed and management without needing a local storage infrastructure [Wan+13]. The centralised architecture of cloud storage depends on the use case and the data protection characteristics for the data that is stored. One would for example consider the physical and technical storage connections based on costs or performance. A Serial AT Attachment (SATA) connection would be preferable to keep costs down and the Non-Volatile Memory Express (NVMe) would increase performance, but also costs. Uses cases that drive this decision could be distinguishing between day-to-day used cloud storage (need for performance) or an irregular used long-term archive storage (cost optimization) [Xu+15]. Another characteristic of cloud storage is the way data is accessed. Two common ways are either using Application Programming Interfaces (APIs) to integrate data transfer directly into applications or file bases protocols for data and file transfer independent of applications (e.g. File Transfer Protocol) [KH20; Lis21]. For the emerging trends discussed in this report, the following cloud characteristics are most relevant:

Edge Computing

- **Broad Network Access**, for seamless interaction between diverse edge devices and central cloud storage. APIs are important to consider for this case due to the wide range of heterogeneous applications and systems [XJK19].
- **On-demand Self Service**, as it enables devices at the edge to access resources, as needed, for real-time processing, also requiring fast access speeds.

Sustainability in Cloud

- **Rapid Elasticity**, to scale resource usage based on demand and therefore save resources if not needed.
- **Measured Services**, for transparent monitoring of cloud resource usage for efficient utilization.
- **Physical and technical storage connections**, needs to be considered for the use case of the storage system, as SATA drives are more energy efficient in comparison to NVMe connections [Wil24].

AI in Data Management

- **Resource Pooling**, for efficient allocation of computing resources for AI systems engaged in metadata management, governance, and more.
- **Broad Network Access**, to ensure the availability of AI assisted data management on heterogeneous systems.
- **Access Protocols**, like APIs are crucial for AI in data management, as they allow seamless communication between the storage, systems and AI applications [AIK20].

3 Edge Computing

Edge computing differentiates itself from traditional cloud computing, as computing tasks are performed de-centralised, closer to data devices (i.e. edge devices) in comparison to traditional centralised storage and processing. Computing tasks and therefore cloud services are generally pushed from the network core to the network edge [Sat17]. Resources are unified closer to the user to provide resources with close proximity to the network [Zha+18]. With the emergence of Internet of Things (IoT) devices, more and more gadgets become data creators and consumers at the network edge [Zho+19]. This reduces the latency of the cloud which is also relevant for different use cases that will further be discussed in this report [Sim+21]. In general, edge computing is an architecture that is complementary to cloud computing and leverages the already existing cloud components [Red22]. Figure 2 provides a simplified illustration regarding the communication paths within the network of the cloud and edge devices.

3.1 Edge Computing as a Trend

To highlight the importance of edge computing as a trend in the realm of cloud storage, the current technological advancements in IoT devices need to be considered. This trend led to an exponential increase in internet traffic in last years. Some applications that drive this increased traffic are for example online gaming and ultra high definition video streaming. This is further stimulated through the wide-spread expansion of the 5G network that increases wireless mobile speeds [Mao+17; Cao+20; LXZ15]. The predicted increase in data created and consumed over the coming years therefore demands a reduction of latency for data transmission and high access speeds [Mao+17; Sta23b]. For optimal local consumption of the increasing data volume, de-centralised storage and caching is necessary [Mao+17; Cao+20]. From an economical point of view, this will also drive market size of edge computing (predicted to be 317 billion USD in revenue by 2026) [Sta23a]. Edge computing is therefore a trend with a lot of gravity in the domain of cloud storage.

3.2 Edge Computing Functionality

In a conceptual view, edge computing can be represented with three layers: (1) cloud layer, (2) edge computing layer and the (3) edge device layer (See figure 2 for a visualisation of the architecture) [STP22]. The cloud layer is represented by traditional, central cloud infrastructure (e.g. data centers) with storage facilities. These can be deployed globally and are also used for large scale processing tasks and long-term storage [Cao+20; Lin+20]. Edge computing is characterized by bringing storage and computing resources closer to the edge of the network (i.e. more local) which is depicted by the de-centralised edge computing and edge device layers [STP22]. Between the edge devices and the central cloud lay the edge nodes. These can vary in size and for example be micro data centers, or storage resources connected to a router [Zho+19]. Essentially, these nodes provide additional computing and storage resources, with low latency as the data does not need to be transferred to cloud storages [HV19]. As described before, IoT devices represent the dasta consumption and production part of the edge device layer. These devices can also request services from the cloud and perform computing tasks, like offloading, caching and processing. One common edge devices are smartphones, that are additionally equipped with sensors [Shi+16; Zha+19].

There are also industry standards that have been established for the architecture e.g., by the Edge Computing Consortium (ECC). With their edge computing reference frame 3.0 more details are provided from different point of views. Different services, like security services, data life cycle services, and management services (e.g, monitoring or data analysis) are portrayed to span across all three edge computing layers. The framework also consists of deployment strategies for distribution, also including APIs for local access [Cao+20]. At the time of writing this report, the website of the ECC has not been accessible, making it uncertain if this reference frame 3.0 is still an industry standard.

Another framework is the EdgeX Foundry, which is open source and hosted by the Linux Foundation. It describes a "northbound" (e.g. cloud data centers) and "southbound" (e.g. edge devices) area that represent the outer most layers of edge computing which communicate with the middle layer, the edge network. Like the reference frame 3.0 they also describe different services, but they mainly reside in the network layer, not across all edge layers. One exemplary service layer are the core services, which can be connected to through APIs and offer services like metadata management. The main use

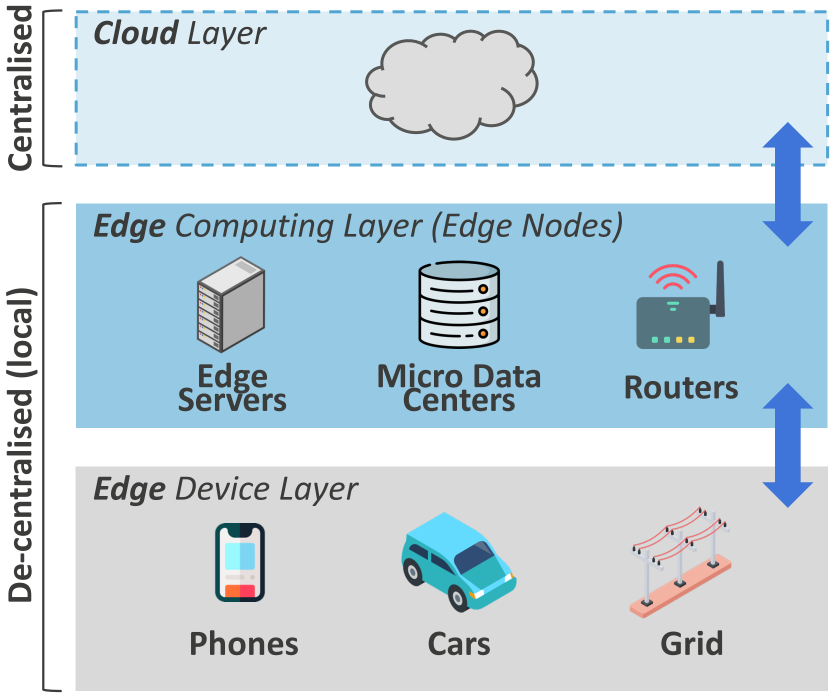

Figure 2: Reference architecture, own illustration based on Sepulveda et al. [STP22].

case of EdgeX Foundry are industrial IoT applications. Contrary to the reference frame 3.0 the EdgeX Foundry is a widely adapted framework with presented use cases from Intel, Hewlett Packard, and Accenture, among other companies [Edg24; Cao+20].

3.2.1 Edge Computing for Media Streaming

Video streaming is a use case example that exemplifies a need for large amounts of bandwidth, especially for live streaming, like on websites like www.twitch.tv or www.youtube.com. The content delivery in this case oftentimes happens as data creation and consumption on smartphones as edge devices. With today's capabilities, they offer the possibility for data processing outside of central cloud locations. Data only needs to reach cloud facilities for long-term storage. Transcoding and processing of video data can happen on the edge devices themselves (i.e. smartphones) [BE17; Ami+23]. Researchers further optimized this approach for higher consumer demands and more heterogeneous systems by deploying light weight transcoding at the edge. This approach essentially stores metadata about the most popular video segments. This information is used in edge servers to store the popular segments in all available bit rates. This includes lower bit rates for fast transmission in low bandwidth areas. Unpopular video parts are only stored in the highest bit rate. This reduces transcoding time and efficiency on the edge, as only unpopular video segments need additional transcoding from devices [Erf+21].

3.2.2 Edge Computing for Autonomous Vehicles

Edge computing is essential for autonomous vehicles, as they rely on data processing from heterogeneous sources (e.g., wide variety of vehicle sensors) with as little latency as possible, due to potentially high speeds on the road and the necessity for fast reactions. The sensory data that is processed is, among other use cases, applied for smart navigation or obstacle detection. This means that all kind of different technologies are at play, like distance sensing and computer vision. Edge computing brings the processing power into the vehicle hardware for reduced latency and increased efficiency. One also needs to keep in mind that high bandwidth network connections are not available everywhere on the road [Liu+19]. Chen et al. developed a framework called F-Cooper which deploys a co-operative approach of data sharing and processing between vehicles. The communication between vehicles is only feature based, without sending the full raw data to avoid network congestion and reduce processing speed [Che+19].

Further developments in edge computing also encourage the inclusion of AI: This technology could be leveraged to analyse data with regards to the proximity to the data source for efficient processing. Also, the large amounts of data produced and consumed open the gate for future machine learning use cases to improve decision making in various fields [Xu+20]. The ongoing democratization of AI is another aspect that highlights the synergy between those two concepts. There can be different degrees of division of work between the edge devices and cloud resources where either most of the AI training happens in the cloud up to having everything be done on edge devices. The merging of AI and edge computing is also known as edge intelligence [Zho+19].

4 Sustainability in Cloud

Sustainability in cloud in this report focuses on sustainability in an environmental context (in contrast to social or economical sustainability). The goal is to reduce the environmental impact of cloud services and also promote responsible resource usage. Efficient management of energy consumption or data processing in cloud computing are therefore key in this frame of reference. [Bal+11; Pen+19]. Accordingly, the ultimate goal is to reduce the impact of the cloud on climate change.

4.1 Sustainability in Cloud as a Trend

As the world's awareness towards environmental sustainability increases, cloud computing offers mechanisms that can help it become more environmental friendly. Large corporations are already leveraging the opportunities that cloud provides regarding the reduction of climate impact:

> "As of 2020, Facebook's operations are now supported by 100 % renewable energy and have reached net zero emissions." [RF21]

> "DeepMind AI Reduces Google Data Centre Cooling Bill by 40 %." [EG16]

Furthermore, Statista reports a green data center market value in Europe of 14.57 billion USD in 2023. They also predict this to increase up to 49.84 billion USD by 2030 which is an increase by almost 350 % [Sta23c]. Also, from the beginning of 2023, companies are required to report on their sustainability as part of the corporate sustainability reporting directive by the European Union. The energy consumption in data centers is also predicted to rise 15 - 20 % each year [MSK13]. In addition to the evident climate impact reduction, this gives companies an incentive to increase the priority of sustainability in cloud for their day to day business [Eur23]. As cloud storage represents a major hardware part of cloud computing, it can be used as the main lever to increase environmental sustainability. The following section provides an overview into different sustainability indicators and how to measure them.

4.2 Towards Measurable Sustainability in Cloud Storage

Due to the breadth of the environmental sustainability term, it can be overwhelming to find sufficient ways of decreasing the climate impact of cloud storage. Storage capacities of data centers are also expected to increase 50 % - 70 % annually [Shu+16]. In comparison to processors, memory, network and cooling, about 20 % of energy in data centers is consumed by storage [GB19]. Gill and Buyya developed a taxonomy that consists of a range of metrics that help quantify sustainability measures for cloud data centers (see figure 3) [GB19].

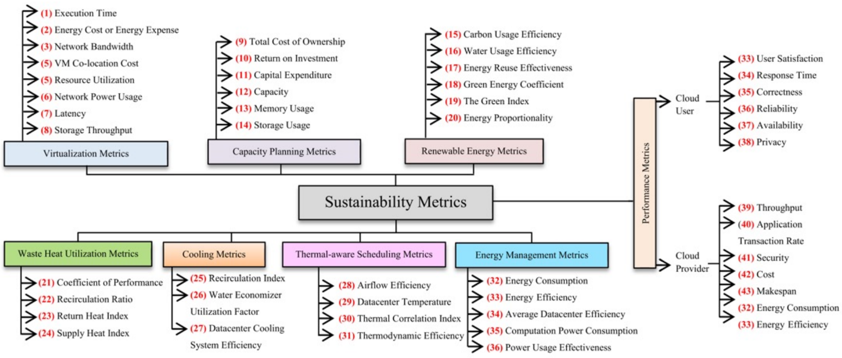

Figure 3: Sustainability metrics taxonomy developed by Gill and Buyya [GB19].

Using renewable energy as an example: There is an increasing public interest in renewable energy, as well as reducing the consumption of energy. Using wind, solar or water energy are broad measures to increase the use of renewable energy. The green energy coefficient, defined as the "ratio of green energy consumed by a cloud data center to total energy consumption of that cloud data center" is one metric that enables cloud providers to quantify how much renewables are involved in their infrastructure [GB19]. The following segments present different metric categories that also provide specific measures related to storage.

Capacity Planning Metrics

The goal of capacity planning is to maximize resource usage. The capacity should be planned for various components of a cloud data center, like information technology devices (e.g. storage), power infrastructure or cooling. Recovery mechanisms, as well as general utilization need to be considered to maximise customer satisfaction when planning for cloud storage [GB19]. The trade-off between low service downtime costs and high operational costs needs to be managed and determined to establish an optimal number of physical machines to guarantee smooth operation. Ghosh et al. developed an algorithm to calculate an optimal amount of physical machines that minimizes the cost of ownership including operational cost with a predetermined service level agreement for downtime requirements. This generally helps deciding between cheaper and less reliable components in comparison to costlier but more reliable machines for the same service level agreement [Gho+14]. One metric that is particularly of interest for cloud storage is storage usage. It is defined as "the total usage of secondary memory (hard disk) to execute user tasks"

[GB19]. By managing storage resources efficiently, energy usage and electronic waste can be reduced to increase environmental sustainability.

Virtualization Metrics

Virtualization for cloud hardware is another technology that increases efficiency and also sustainability. The provisioning of cloud resources to customers can be more dynamic to adhere to user's needs and reduce energy usage, if not needed [Gia+15]. The technology enables the creation of a virtual version on physical existing computing resources in a cloud environment. This allows multiple virtual machines to operate on a single physical machine to also reduce the physical space needed in a data center. Hardware gets abstracted into software for easier remote management [Ama23]. Some metrics that are relevant for a greener cloud in regards to virtual machines are for example load balancing, scheduling, or fault tolerance [Wan+17]. In this context, storage throughput is defined as "the amount of time taken for the storage system to execute the required operation per second". This implies, that the faster the execution on a computing task, the smaller the environmental impact due to lower energy consumption [GB19; HZP20].

4.3 Related Green Cloud Strategies

After highlighting different metrics related to cloud storage that help measure environmental impact, this section presents different strategies that help move towards greener cloud. The GreenPlanning framework highlights the importance of diversifying energy sources and bringing more green sources into the portfolio to reduce emissions, by helping to choose the fitting sources and also incorporating energy storage systems, thereby reducing the capital and operational costs. To keep service availability as high as possible, the framework proposes to use Energy Storage Devices (ESD) that mitigate the unreliability of renewable energy sources like wind or solar (see figure 4).([KL16]

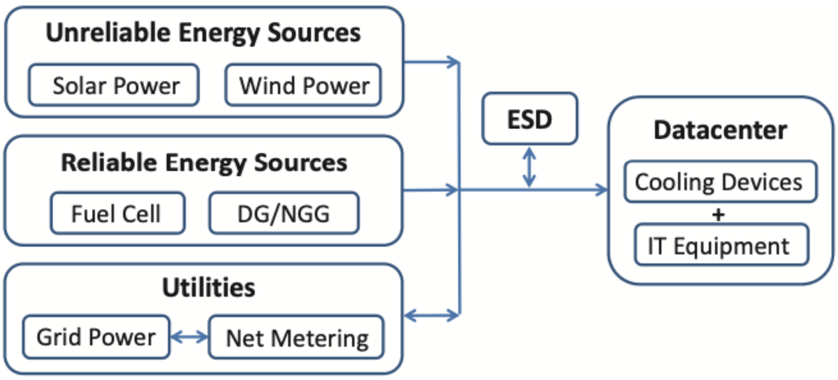

Figure 4: Exemplary data center structure with multiple energy sources [KL16].

Another example in choosing cloud infrastructure (i.e. data centers) in a distributed cloud environment to minimize the carbon footprint: Chamkoori and Katebi found a way to reduce the distance between data centers for a data set up to 30 % by deploying a particle swarm optimization algorithm that improves the selection of storage service based on their physical location [CK23].

Luo et al. developed a system called Boafft that uses distributed deduplication to increase storage efficiency in big data applications. This technology scans data in distributed systems scanned and analyses it to find duplicates. This reduces the amount of duplicate data and therefore the overall storage footprint to reduce energy usage [Luo+20].

With a focus on energy consumption, companies should choose the source for their energy carefully and complementary to their physical location. Microsoft for example chose to build data centers near dams to harness the hydro-power generated by water flow. It is also possible to use waterways as means of transport to reduce the carbon emissions even before a data center has been built [SC21].

5 AI in Data Management

With an increase of 28 % annually in data being created, captured and consumed, managing said data becomes more complex each year [Sta23b]. Data management consists of a range of tasks that need to be fulfilled, such as data governance, security, quality or metadata management that are related to storage in cloud systems [tab]. AI in data management is an important trend that currently transforms the practice and assist in tasks such as data extraction, mapping, quality control and analysis. It is able to extract unstructured data from heterogeneous sources and offers the possibility to automate otherwise time consuming tasks. Other aspects of AI can also be assisting with the discovery of data sources and relationships between data, as well as increasing data quality by automatically identifying errors in patterns of data. Organizations need to consider this technology to efficiently scale their data management in the cloud [Dat23].

5.1 AI in Metadata Management

In simple terms, metadata can be described as the data about data and makes it descriptive and structures it [PJ22]. Metadata management therefore revolves around identifying labels for data, classifying data and is usually a time consuming task, when done manually [Exp22]. A practical example has been presented by Harper et al. about research focused libraries that need AI support to organize scientific texts, such as electronic theses and dissertations. They describe the traditional vocabulary libraries that are used to generate metadata but are increasingly more time consuming with the rise of digitally created texts. Also, keywords provided by authors tend to produce less accurate search results with the growing amount of theses created. To address this challenge and keep high quality search results the authors deployed an unsupervised machine learning model called Doc2Vec. The system transforms text into numerical vectors to understand the content and context of a publication to then tag it with a large, structured data set derived from Wikipedia as a source for keywords. They further deployed a latent dirichlet allocation model that clusters related words and creates "topics" based on them. The authors found an increase in discoverability but recommend also using human expertise in conjunction to an unsupervised learning model [Har+23].

Amazon Web Services (AWS) also provides a service called DataZone that uses a machine learning model to generate metadata and increase discoverability of data. Additionally, they provide services that support with data governance [Roa23].

5.2 AI in Data Governance

Data governance is needed for an organization to manager their data effectively. It involves processes, roles (e.g. Data Owner, Data Stewards, etc.), policies, metrics and standards that ensure efficient data usage throughout its life cycle. Other goals are also improving decision making or ensuring compliance to regulations. This implies that data quality and security of data play an important role. As data governance implies a wide range of activities, data management cannot exist without it [Mic23; IBM24]. Martin et al. propose different techniques that leverage AI to increase data quality in the context of IoT, specifically air-quality monitoring. They argue that one of the most important quality indicators is "fitness for use", in addition to a objective multi-dimensional approach. The dimension in use are namely: accuracy, completeness, timeliness, precision, and usability. For an AI system to properly use this dimensional approach, they provide calculation methods for each dimension. The study used 100 sensors that created about 10,000 data items. They then curated the data using their domain knowledge and created a high-quality data set for statistical training. They then trained and deployed several mechanisms to increase data quality in their sensor data: (1) Outlier detection by using an Isolation Forest and a Local Outlier Factor model to detect anomalies in the data. (2) Assigning missing values by using a Polynomial Interpolation of Degree 2 and k-Nearest Neighbors model. (3) Novelty detection by again deploying an Isolation forest, Local Outlier Factor and One-Class SVM model. At last, they also used the (4) Seasonal AutoRegressive Integrated Moving Average algorithm to extend the data set with synthetic data. They reported a positive conclusion in deploying AI as a viable strategy to increase data quality. An important limitation though is the single use case with homogeneous data they used which might make the application more difficult in a more heterogeneous environment [Mar+23].

5.3 Further Data Management Applications

Other noteworthy applications of AI in data management are for example "Constance", a system that provides intelligent data management in data lakes, usually schema-less "dump" for raw data. The system is able to discover, extract and summarise data to develop and assign metadata to it [HGQ16].

Data fabric is an approach that can also leverage AI in its application. The goal is to harmonize heterogeneous data sources on-premise, as well as in the cloud and unify them. This enables seamless access to data, as well as quality controls and governance mechanisms. In organisations it can also help break down data silos and increase transparency across departments. The technologies available to data fabric also depend on the specific architecture and provider that implements it. In a scientific context, this architecture has already been studied in industrial applications for road transport systems. [IBM23; Kuf+23]

6 Conclusion

The report presents insights into three chosen trends in cloud storage, namely edge computing, sustainability in cloud and AI in data management. With the use cases described, the importance for real world applications is emphasized. These three areas should not

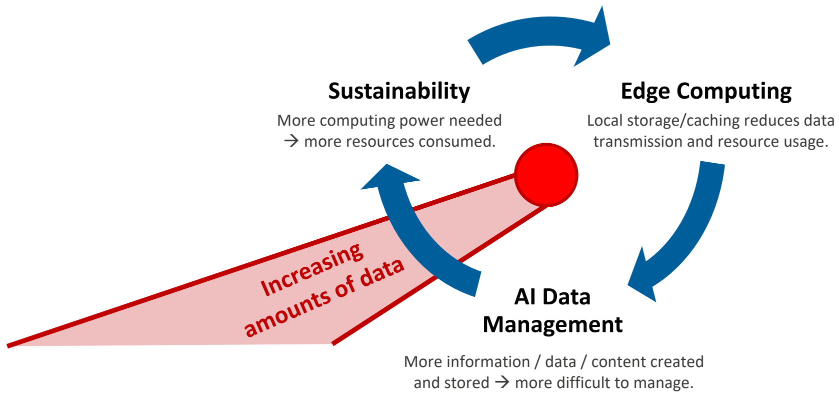

Figure 5: Relation between trends

be viewed as different silos, but as deeply interwoven that share concepts and build upon each other. The shift towards edge computing shows how cloud evolved from central data storage to highly complex and efficiency driven distributed systems. The main drivers are the need for low latency, the increase in IoT devices and need for real-time data processing. Sustainability has become a new imperative in our society, as well as in cloud storage. This growing recognition is also supported by large, publicly known corporations that help drive the transition towards lower environmental impact of cloud infrastructure. Central levers are the utilization of renewable energy sources, as well as increased technical efficiency that align organizational strategical goals and cost reductions with lower carbon emissions. Lastly, AI technologies make their way into data management applications which will become more essential with the increasing amounts of data produced each year. Automating complex tasks, like data quality measures and metadata management makes AI a pivotal force for the future by increasing scalability and efficiency. As depicted in 5, the three trends mainly revolve around the increasing amount of data and are in a cycle relation to each other: The more data and demand for data, the more computing power needed on the edge of the network, which also drives the need for efficient and automatic management of the data. With the need for more cloud infrastructure, more resources are consumed which drives the need for sustainability measures in cloud storage. A future report could also highlight the use of AI technology to create sustainable hardware architectures, or improve demand-driven storage utilization. Another aspect would be the exploration of more general applicable AI data management cases, as the currently highlighted are reliant on specific domain applications. Another aspect to consider could also be increasing data security with AI systems. Expanding the sustainability topic is also important, as the socio-economic changes due to increased data need to be emphasized. This could include data literacy and data protection awareness for users for example.

References

[AIK20] M. Ariffin, M. Ibrahim, and Z. Kasiran. "Api vulnerabilities in cloud computing platform: attack and detection". In: *International Journal of Engineering Trends and Technology* (2020), pp. 8–14. DOI: 10.14445/22315381/cati1p202.

[Ama23] Amazon Web Services. *What is Virtualization? - Cloud Computing Virtualization Explained*. https://aws.amazon.com/what-is/virtualization/. Accessed: 2024-03-31. 2023.

[Ami+23] H. Amirpour et al. "Cd-lwte: cost- and delay-aware light-weight transcoding at the edge". In: *IEEE Transactions on Network and Service Management* 20 (3 2023), pp. 3104–3118. DOI: 10.1109/tnsm.2022.3229744.

[Bal+11] J Baliga et al. "Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport". In: *Proceedings of the IEEE* 99.1 (Jan. 2011), pp. 149–167. ISSN: 0018-9219, 1558-2256. DOI: 10.1109/JPROC.2010.2060451. URL: http://ieeexplore.ieee.org/document/5559320/ (visited on 12/14/2023).

[BE17] Kashif Bilal and Aiman Erbad. "Edge computing for interactive media and video streaming". In: *2017 Second International Conference on Fog and Mobile Edge Computing (FMEC)*. Valencia, Spain: IEEE, May 2017, pp. 68–73. ISBN: 978-1-5386-2859-1. DOI: 10.1109/FMEC.2017.7946410. URL: http://ieeexplore.ieee.org/document/7946410/ (visited on 12/07/2023).

[Cao+20] Keyan Cao et al. "An Overview on Edge Computing Research". In: *IEEE Access* 8 (2020), pp. 85714–85728. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2020.2991734. URL: https://ieeexplore.ieee.org/document/9083958/ (visited on 03/23/2024).

[Che+19] Qi Chen et al. "F-cooper: feature based cooperative perception for autonomous vehicle edge computing system using 3D point clouds". In: *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*. SEC '19. Arlington, Virginia: Association for Computing Machinery, 2019, pp. 88–100. ISBN: 9781450367332. DOI: 10.1145/3318216.3363300. URL: https://doi.org/10.1145/3318216.3363300.

[CK23] Alireza Chamkoori and Serajdean Katebi. "Security and storage improvement in distributed cloud data centers by increasing reliability based on particle swarm optimization and artificial immune system algorithms". In: *Concurrency and Computation: Practice and Experience* 35.6 (2023), e7580. DOI: https://doi.org/10.1002/cpe.7580. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.7580. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.7580.

[Dat23] DataCamp. *How is AI Transforming Data Management? Revolutionizing Extraction, Mapping & Analysis*. Accessed: 2024-03-31. 2023. URL: https://www.datacamp.com/community/blog/ai-transforming-data-management.

[Edg24] EdgeX Foundry Project. *EdgeX Foundry | The Open Source Edge Platform*. https://www.edgexfoundry.org. Accessed: 2024-03-24. LF Projects, LLC, 2024.

[EG16] Richard Evans and Jim Gao. *DeepMind AI Reduces Google Data Centre Cooling Bill by 40%*. 2016. URL: https://deepmind.google/discover/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40/.

[Erf+21] Alireza Erfanian et al. "LwTE: Light-Weight Transcoding at the Edge". In: *IEEE Access* 9 (2021), pp. 112276–112289. DOI: 10.1109/ACCESS.2021.3102633.

[Eur23] European Commission. *Corporate Sustainability Reporting*. https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en. Accessed: 2024-03-25. 2023.

[Exp22] Expert.ai. *What Is Metadata Management and Why Is It Important?* Tech. rep. 2022. URL: https://www.expert.ai/blog/what-is-metadata-management-and-why-is-it-important/.

[GB19] Sukhpal Singh Gill and Rajkumar Buyya. "A Taxonomy and Future Directions for Sustainable Cloud Computing: 360 Degree View". en. In: *ACM Computing Surveys* 51.5 (Sept. 2019), pp. 1–33. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3241038. URL: https://dl.acm.org/doi/10.1145/3241038 (visited on 12/14/2023).

[Gho+14] Rahul Ghosh et al. "Stochastic Model Driven Capacity Planning for an Infrastructure-as-a-Service Cloud". In: *IEEE Transactions on Services Computing* 7.4 (Oct. 2014), pp. 667–680. ISSN: 1939-1374. DOI: 10.1109/TSC.2013.44. URL: http://ieeexplore.ieee.org/document/6594736/ (visited on 01/10/2024).

[Gia+15] Maurizio Giacobbe et al. "An approach to reduce carbon dioxide emissions through virtual machine migrations in a sustainable cloud federation". In: *2015 Sustainable Internet and ICT for Sustainability (SustainIT)*. 2015, pp. 1–4. DOI: 10.1109/SustainIT.2015.7101383.

[Har+23] Charlie Harper et al. *Using AI for Metadata Tagging to Improve Resource Discovery*. https://www.choice360.org/libtech-insight/using-ai-for-metadata-tagging-to-improve-resource-discovery/. Accessed: 2024-03-31. Apr. 2023.

[HGQ16] Rihan Hai, Sandra Geisler, and Christoph Quix. "Constance: An Intelligent Data Lake System". en. In: *Proceedings of the 2016 International Conference on Management of Data*. San Francisco California USA: ACM, June 2016, pp. 2097–2100. ISBN: 978-1-4503-3531-7. DOI: 10.1145/2882903.2899389. URL: https://dl.acm.org/doi/10.1145/2882903.2899389 (visited on 03/31/2024).

[HV19] C. Hong and B. Varghese. "Resource management in fog/edge computing". In: *ACM Computing Surveys* 52 (5 2019), pp. 1–37. DOI: 10.1145/3326066.

[HZP20] H. He, Y. Zhao, and S. Pang. "Stochastic modeling and performance analysis of energy-aware cloud data center based on dynamic scalable stochastic petri net". In: *Computing and Informatics* 39 (1-2 2020), pp. 28–50. DOI: 10.31577/cai_2020_1-2_28.

[IBM23] IBM. *What is a Data Fabric?* https://www.ibm.com/topics/data-fabric. Accessed: 2024-03-31. 2023.

[IBM24] IBM. *What is Data Governance?* https://www.ibm.com/topics/data-governance. Accessed: 2024-03-31. 2024.

[IEE] IEEE. *Real-Life Use Cases for Edge Computing.* URL: https://innovationatwork.ieee.org/real-life-edge-computing-use-cases/ (visited on 02/01/2024).

[KH20] Daniel Kirsch and Judith Hurwitz. *Cloud computing for dummies.* 2nd. Indianapolis: John Wiley and Sons, 2020. ISBN: 978-1-119-54665-8.

[KL16] Fanxin Kong and Xue Liu. "GreenPlanning: Optimal Energy Source Selection and Capacity Planning for Green Datacenters". In: *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS).* 2016, pp. 1–10. DOI: 10.1109/ICCPS.2016.7479104.

[Kuf+23] N. G. Kuftinova et al. "Implementation of the Data Fabric Architecture as a Sustainable Development of Industrial Platform Technologies in Road Transport Systems". In: *2023 Systems of Signals Generating and Processing in the Field of on Board Communications.* Moscow, Russian Federation: IEEE, Mar. 2023, pp. 1–5. ISBN: 9798350397727. DOI: 10.1109/IEEECONF56737.2023.10092048. URL: https://ieeexplore.ieee.org/document/10092048/ (visited on 03/31/2024).

[Lin+20] B. Lin et al. "Cost-driven off-loading for dnn-based applications over cloud, edge, and end devices". In: *Ieee Transactions on Industrial Informatics* 16 (8 2020), pp. 5456–5466. DOI: 10.1109/tii.2019.2961237.

[Lis21] Anders Lisdorf. *Cloud Computing Basics: A Non-Technical Introduction.* en. Berkeley, CA: Apress, 2021. ISBN: 978-1-4842-6920-6 978-1-4842-6921-3. DOI: 10.1007/978-1-4842-6921-3. URL: https://link.springer.com/10.1007/978-1-4842-6921-3 (visited on 03/13/2024).

[Liu+19] Shaoshan Liu et al. "Edge Computing for Autonomous Driving: Opportunities and Challenges". In: *Proceedings of the IEEE* 107.8 (Aug. 2019), pp. 1697–1716. ISSN: 0018-9219, 1558-2256. DOI: 10.1109/JPROC.2019.2915983. URL: https://ieeexplore.ieee.org/document/8744265/ (visited on 12/07/2023).

[Luo+20] Shengmei Luo et al. "Boafft: Distributed Deduplication for Big Data Storage in the Cloud". In: *IEEE Transactions on Cloud Computing* 8.4 (2020), pp. 1199–1211. DOI: 10.1109/TCC.2015.2511752.

[LXZ15] Shancang Li, Li Da Xu, and Shanshan Zhao. "The internet of things: a survey". en. In: *Information Systems Frontiers* 17.2 (Apr. 2015), pp. 243–259. ISSN: 1387-3326, 1572-9419. DOI: 10.1007/s10796-014-9492-7. URL: http://link.springer.com/10.1007/s10796-014-9492-7 (visited on 03/23/2024).

[Mao+17] Yuyi Mao et al. "A Survey on Mobile Edge Computing: The Communication Perspective". In: *IEEE Communications Surveys & Tutorials* 19.4 (2017), pp. 2322–2358. ISSN: 1553-877X. DOI: 10.1109/COMST.2017.2745201. URL: http://ieeexplore.ieee.org/document/8016573/ (visited on 12/07/2023).

[Mar+23] Laura Martín et al. "Development and evaluation of Artificial Intelligence techniques for IoT data quality assessment and curation". en. In: *Internet of Things* 22 (July 2023), p. 100779. ISSN: 25426605. DOI: 10.1016/j.iot.2023.100779. URL: https://linkinghub.elsevier.com/retrieve/pii/S2542660523001026 (visited on 03/31/2024).

[MGN] Peter Mell, Timothy Grance, and National Institute of Standards and Technology. *The NIST Definition of Cloud Computing*. Tech. rep. URL: https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-145.pdf.

[Mic23] Microsoft Azure. *What is Data Governance?* https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-a-data-governance/. Accessed: 2024-03-31. 2023.

[MSK13] Eric Masanet, Arman Shehabi, and Jonathan Koomey. "Characteristics of low-carbon data centres". en. In: *Nature Climate Change* 3.7 (July 2013), pp. 627–630. ISSN: 1758-678X, 1758-6798. DOI: 10.1038/nclimate1786. URL: https://www.nature.com/articles/nclimate1786 (visited on 03/25/2024).

[Pen+19] Keith L. Pendergrass et al. "Toward Environmentally Sustainable Digital Preservation". en. In: *The American Archivist* 82.1 (Mar. 2019), pp. 165–206. ISSN: 0360-9081. DOI: 10.17723/0360-9081-82.1.165. URL: http://americanarchivist.org/doi/10.17723/0360-9081-82.1.165 (visited on 12/14/2023).

[PJ22] Jaihyun Park and Sullam Jeoung. "Raison d'être of the benchmark dataset: A Survey of Current Practices of Benchmark Dataset Sharing Platforms". en. In: *Proceedings of NLP Power! The First Workshop on Efficient Benchmarking in NLP*. Dublin, Ireland: Association for Computational Linguistics, 2022, pp. 1–10. DOI: 10.18653/v1/2022.nlppower-1.1. URL: https://aclanthology.org/2022.nlppower-1.1 (visited on 01/11/2024).

[Red22] Red Hat. *What is edge architecture?* Mar. 2022. URL: https://www.redhat.com/en/topics/edge-computing/what-is-edge-architecture (visited on 07/12/2023).

[RF21] Melanie Roe and Facebook. *Facebook Reaches 100% Renewable Energy*. Tech. rep. 2021. URL: https://sustainability.fb.com/wp-content/uploads/2021/04/Facebook_RenewableEnergy_April2021.pdf.

[Roa23] John Roach. *How Amazon DataZone helps customers find value in oceans of data*. https://www.amazon.science/news-and-features/how-amazon-datazone-helps-with-data-management-automated-metadata-generation. Accessed: 2024-03-31. 2023.

[Sat17] Mahadev Satyanarayanan. "The Emergence of Edge Computing". In: *Computer* 50.1 (2017), pp. 30–39. DOI: 10.1109/MC.2017.9.

[SC21] Dharshika S and Nagaraj G Cholli. "Green Cloud Computing: Redefining the future of Cloud Computing". en. In: *International Research Journal on Advanced Science Hub* 3.Special Issue 7S (July 2021), pp. 12–19. ISSN: 2582-4376. DOI: 10.47392/irjash.2021.203. URL: https://rspsciencehub.com/article_12114.html (visited on 12/14/2023).

[Shi+16] Weisong Shi et al. "Edge Computing: Vision and Challenges". In: *IEEE Internet of Things Journal* 3.5 (Oct. 2016), pp. 637–646. ISSN: 2327-4662. DOI: 10.1109/JIOT.2016.2579198. URL: http://ieeexplore.ieee.org/document/7488250/ (visited on 12/07/2023).

[Shu+16] Junaid Shuja et al. "Survey of Techniques and Architectures for Designing Energy-Efficient Data Centers". In: *IEEE Systems Journal* 10.2 (June 2016), pp. 507–519. ISSN: 1932-8184, 1937-9234, 2373-7816. DOI: 10.1109/JSYST.2014.2315823. URL: https://ieeexplore.ieee.org/document/6856153 (visited on 03/25/2024).

[Sim+21] Milos Simic et al. "Towards Edge Computing as a Service: Dynamic Formation of the Micro Data-Centers". In: *IEEE Access* 9 (2021), pp. 114468–114484. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2021.3104475. URL: https://ieeexplore.ieee.org/document/9512038/ (visited on 12/06/2023).

[Sta23a] Statista. *Internet of Things: market data & analysis*. Tech. rep. 2023. URL: https://www.statista.com/study/109197/internet-of-things-market-outlook-report/.

[Sta23b] Statista. *Public Cloud: market data & analysis*. Tech. rep. 2023. URL: https://www.statista.com/study/85676/public-cloud-report/.

[Sta23c] Statista. *Value of the green data center market in Europe from 2020 to 2023, with a forecast until 2030*. Tech. rep. 2023. URL: https://www.statista.com/statistics/1398530/europe-green-data-center-market/.

[STP22] Frank Sepulveda, Joseph Soloman Thangraj, and Jay Pulliam. "The Edge of Exploration: An Edge Storage and Computing Framework for Ambient Noise Seismic Interferometry Using Internet of Things Based Sensor Networks". en. In: *Sensors* 22.10 (May 2022), p. 3615. ISSN: 1424-8220. DOI: 10.3390/s22103615. URL: https://www.mdpi.com/1424-8220/22/10/3615 (visited on 12/07/2023).

[tab] tableau. *Data Management: What It Is, Importance, And Challenges*. URL: https://www.tableau.com/learn/articles/what-is-data-management.

[TOM16] Y. Tarutani, Y. Ohsita, and M. Murata. "Placement of virtual storages for distributed robust cloud storage". In: *IEICE Transactions on Communications* E99.B (4 2016), pp. 885–893. DOI: 10.1587/transcom.2015ebp3292.

[Wan+13] Cong Wang et al. "Privacy-Preserving Public Auditing for Secure Cloud Storage". In: *IEEE Transactions on Computers* 62.2 (Feb. 2013), pp. 362–375. ISSN: 0018-9340. DOI: 10.1109/TC.2011.245. URL: http://ieeexplore.ieee.org/document/6109245/ (visited on 01/01/2024).

[Wan+17] Xiaoying Wang et al. "Green-Aware Virtual Machine Migration Strategy in Sustainable Cloud Computing Environments". en. In: *Cloud Computing - Architecture and Applications*. Ed. by Jaydip Sen. InTech, June 2017. ISBN: 978-953-51-3243-1 978-953-51-3244-8. DOI: 10.5772/67350. URL: http://www.intechopen.com/books/cloud-computing-architecture-and-applications/green-aware-virtual-machine-migration-strategy-in-sustainable-cloud-computing-environments (visited on 03/31/2024).

[Wil24] Harold Williams. *Nvme vs SATA SSD power consumption: Which one is better?* Mar. 2024. URL: `https://driveshero.com/nvme-vs-sata-ssd-power-consumption/`.

[XJK19] R. Xu, W. Jin, and D. Kim. "Microservice security agent based on api gateway in edge computing". In: *Sensors* 19 (22 2019), p. 4905. DOI: `10.3390/s19224905`.

[Xu+15] Qiumin Xu et al. "Performance analysis of NVMe SSDs and their implication on real world databases". In: *Proceedings of the 8th ACM International Systems and Storage Conference* (2015). DOI: `10.1145/2757667.2757684`.

[Xu+20] Dianlei Xu et al. "Edge Intelligence: Architectures, Challenges, and Applications". In: (2020). Publisher: [object Object] Version Number: 2. DOI: `10.48550/ARXIV.2003.12172`. URL: `https://arxiv.org/abs/2003.12172` (visited on 03/19/2024).

[Zha+18] Z. Zhao et al. "Edge Computing: Platforms, Applications and Challenges". In: *Jisuanji Yanjiu yu Fazhan/Computer Research and Development* 55 (Feb. 2018), pp. 327–337. DOI: `10.7544/issn1000-1239.2018.20170228`.

[Zha+19] Yongli Zhao et al. "Edge Computing and Networking: A Survey on Infrastructures and Applications". In: *IEEE Access* 7 (2019), pp. 101213–101230. ISSN: 2169-3536. DOI: `10.1109/ACCESS.2019.2927538`. URL: `https://ieeexplore.ieee.org/document/8758431/` (visited on 12/07/2023).

[Zho+19] Zhi Zhou et al. "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing". In: *Proceedings of the IEEE* 107.8 (Aug. 2019), pp. 1738–1762. ISSN: 0018-9219, 1558-2256. DOI: `10.1109/JPROC.2019.2918951`. URL: `https://ieeexplore.ieee.org/document/8736011/` (visited on 12/06/2023).

[ZLB10] Q. Zhang, C. Lü, and R. Boutaba. "Cloud computing: state-of-the-art and research challenges". In: *Journal of Internet Services and Applications* 1 (1 2010), pp. 7–18. DOI: `10.1007/s13174-010-0007-6`.