Effects of edge computing growth on demand of data centre services

Author: Kristo Lippur
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT,

The increased prevalence and proliferation of edge computing as a paradigm poses both opportunities and challenges for traditional IT infrastructure and cloud data centres. It is impossible to predict the exact effects of these changes in how data is collected, consumed, and used, but enterprises can benefit from taking the potential future scenarios into account when planning their spending and activities. This research analysed three potential scenarios for the effects of edge computing growth on the demand for centralised computing resources for the purposes of processing data created at edge. In the first scenario, the prevalence and therefore the volume of data produced edge computing continues to grow rapidly, with centralised computing demand growing proportionally to aggregate and process the data generated on the edge. In the second scenario, edge computing follows the historic growth pattern of data volume growth, leading to a slower growth in centralised computing demand. In the third scenario, edge computing partially substitutes centralised computing, and many workloads move to the edge, leading to a much smaller increase in centralised computing demand as resources are redirected. In all scenarios, demand for core computing continues to grow, indicating a continued need for investing in new capabilities and additional resources for this purpose.

Graduation Committee members:
Dr. Fons Wijnhoven; Martijn Koot

Keywords
Data centre demand; edge computing; data centres; spending on edge computing data processing; distributed computing.

1. INTRODUCTION

Over the past decade and most notably in the past few years when 5G has become increasingly ubiquitous, with its market share of mobile telecommunication technologies worldwide projected to reach 21% by 2025 (GSMA, 2021), edge computing has gained relevance as an emergent paradigm for distributed computing.

Edge computing as a paradigm means bringing computing resources such as computation and information storage closer to the sources and consumption of data (Khan et al., 2019). Traditionally, computing power has been highly centralised in data centres with hundreds to thousands of servers and hundreds to thousands of gigabits of bandwidth capacity. Edge computing represents a potential change to this status quo by increasing the decentralisation of computing power.

Applications of edge computing as a paradigm include Internet of Things (IoT) devices, smart industry technologies (also known as Industry 4.0), and self-driving cars (vehicular edge computing) (Bilal et al., 2018). In all abovementioned cases, computation largely happens on-site, sometimes in real time, using localised compute resources and data storage facilities. This does not mean isolation from the rest of the network and often connectivity with the outside world is an important value driver for such applications, but data is often pre-processed on the device or near the producer of the data, rather than centrally in a data centre. (Bilal et al., 2018; Khan et al., 2019)

The market for edge computing is growing rapidly, with market revenue in 2021 reaching $176 billion, which is projected to grow to $274 billion by 2025 (IDC & Statista, 2022). Market size for public cloud services is also growing rapidly, with forecasted worldwide market revenues at $396 billion and $482 billion in 2021 and 2022 respectively (Gartner, 2021).

Understanding the role of edge computing infrastructure in the context of this research is important for realistic estimation of the interactions between edge computing and centralised computing. Application of edge computing relies on the availability of sufficient infrastructure to enable processing data in a distributed manner close to the sources and consumers of data. This includes connectivity infrastructure such as mobile networks, including high-speed network technologies such as 5G and fibreoptic networks, as well as reliable electric grids. Such infrastructure can be expensive and complex to deploy and develop, which can hamper application of distributed computing paradigms such as edge computing in areas where the supporting services are not sufficiently developed. These risks and limitations must be taken into account when considering the potential for edge computing to affect centralised computing demand, as lack of sufficient supporting infrastructure can have a negative effect on the volume of data transferred from edge to core. (Khan et al., 2019)

Centralised or core computing (the terms are used interchangeably for the purposes of this research) is defined as computing that takes place in a centralised location such as a data centre, usually on a large scale. This includes virtualised and cloud computing, as the underlying provider of such compute resources is still a physical server usually located in a data centre. Centralised computing is characterised by immobility and large volumes which takes advantage of economies of scale. (Nguyen et al., 2019)

Proximity and low latency have lower priority in the context of centralised computing compared to edge computing, and are usually only considered on a regional scale, with data centres handling the data often located hundreds of kilometres away from the data producer. As such, centralised computing is directly contrasted with edge computing.

Spending on centralised computing has increased by an average of 14% per year between 2018 - 2020. (Synergy Research Group, 2022)

In existing academic literature, significant research has been conducted on the effects and benefits of edge computing to consumers of compute resources, such as businesses who consume cloud services (Khan et al., 2019), as well as changes in data centre energy usage (Koot & Wijnhoven, 2021). However, there is a gap in research addressing the potential effects of the increased prevalence of cloud computing on the providers of cloud services and data centre operators, most notably what the rise in edge computing would mean for future demand for centralised computing resources (large data centres).

The question of whether edge computing is complementary with centralised computing has not been answered. It is therefore not yet clear if the rise in edge computing would increase growth rate of demand for data centres due to the increased need for coordination and communication between different edge computing nodes, or if it would lead to a decrease in the future demand growth for data centres due to distributed computing rendering large data centres obsolete.

This research aims to develop a forecasting model of the impact on the future demand of centralised compute resources by the increasing prevalence of edge computing as an alternative computing paradigm and changes in the usage patterns of data generated from edge devices. The forecasting model will be used to explore a few potential scenarios for future data centre demand based on developments in edge computing.

The main research question is formulated as follows:

How can the growth of edge computing affect the demand of centralised computing resources?

Answering these questions and providing a model for forecasting the demand for data centre services based on the changes in edge computing prevalence would have practical relevance to cloud service providers and data centre operators, as it would enable them to forecast changes in future demand and more accurately prioritise investments and product development. It would also help businesses who provide auxiliary services or develop software for data centres and cloud service providers better predict future trends and development needs. In addition, it is intended to help businesses better forecast expenses related to edge computing adoption within their company.

To support these aims, the research uses spending on core computing resources (data centre services) as the dependent variable and the main result of the forecasting model.

2. METHODOLOGY

The analysis of the potential effects of edge computing on the demand for centralised computing will be done using simulation and scenario analysis based on current and future growth of relevant parameters.

Literature search and review are used to provide the theoretical basis for the research and to identify the potential quantifiable factors describing the relevant effects of edge computing on centralised computing.

The identified factors and the theoretical framework are used to develop various scenarios to answer the research question.

Literature review for this research is conducted using the following data sources:

- Web of Science: used for scientific articles and literature
- Statista: used for statistics and datasets

Web of Science was selected because of their large selection of scientific articles and convenient interface which allows for searching using multiple queries as well as filtering the results by document type and quality indicators such as citations and references.

Statista was selected as it provides access to a large amount of relevant data on the topic and collates information from a large variety of sources. It provides additional background and context for the datasets, which makes them easier to interpret for use in the research.

For developing the theoretical framework for the research proposal, the following queries were used on Web of Science:

- Edge computing
- Edge computing demand
- Cloud computing
- Centralised computing
- Edge computing infrastructure
- Datacentres demand
- Cloud services demand

Each of the queries was used to perform a separate search, including all fields (e.g. "ALL=("edge computing")"). Filtering was applied on the search results to include only articles, in order to filter out editorial materials and other less reliable content. Articles to include were selected from the first 50 results of each search based on the number of citations and references, as well as the relevance to the research topic of the article's title and abstract.

Relevance to the topic of this research was assessed by first selecting articles with titles that clearly addressed the topic under research, after which the selection was narrowed down by reading the article abstracts to exclude research which did not provide relevant information for constructing the theoretical framework of this research, or where the research was too specific (e.g. only focused on a single business sector) to provide general insights needed for this research.

On Statista, the following queries were used:

- Cloud computing market size
- Edge computing
- Datacentre market
- Network infrastructure
- Cloud services demand

Results were selected based on relevance to the research topic by reading the result titles and looking at the data provided. As the result set was relatively small and the data relatively simple, it was possible to assess the relevance by the content of the results.

The results of the literature review form the basis of the forecasting model and scenario analysis developed by this research by providing the definitions and quantification of the model variables, as well as supporting necessary assumptions.

After developing the forecasting model, potential future scenarios for the period of 2020 - 2025 are simulated using different inputs based on historical data and expert forecasts for the variables used in the model.

3. THEORY

For the purposes of this research, the worldwide annual spending on cloud and data centres ("data centre spending" or "core computing spending") will be used as the measure of demand for centralised compute resources. This measure gives the most actionable and relevant information for the intended target audience of this research (decision-makers on IT spending) and is the most reliably quantifiable without excessive assumptions about the future development of computing technologies.

Other ways of quantifying the volume and availability of computing resources were considered for the purposes of this research, such as the availability of CPU, RAM, and digital storage, combined with qualitative indicators, as in research from Nguyen et al. (2019). However, these were not selected due to the lack of publicly available data for these metrics in existing research and statistics, and the collection of this data is outside the scope of this research. In addition, using these indicators would excessively complicate the forecasting model.

Further, quantifying computing demand using resource-specific metrics, such as the amount of CPU, RAM, and digital storage available in data centres, can be misleading, as the cost of compute resources varies over time and is difficult to predict, making the information opaquer. Using such metrics would create an additional burden for the application of the results of this research, as decision-makers would have to predict the future cost of computing resources in order to meaningfully apply the research model. As such, spending can be considered the most actionable and relevant indicator for investment-related decision-making, which best fulfils the goal of this research.

The dependent variable and main result of the model is annual data centre spending for the purpose of processing data created at edge.

3.1 Data transfers from edge to core

In order to make processing and analysis of data produced by edge devices easier and to provide more useful results by aggregating data from a larger pool of edge devices, edge data is being increasingly aggregated using centralised computing resources (the "core"), moving it away from the edge (Figure 1). Seagate found in their research (2020) that only 36% of the 18.2 zettabytes (ZB) of data created on the edge per year was being periodically transferred from edge to core in 2020, but this number was expected to grow to 57% by 2022. They also projected enterprise data collected at the edge to show a compound annual growth rate (CAGR) of 42.2% over the coming years. (Seagate, 2020)

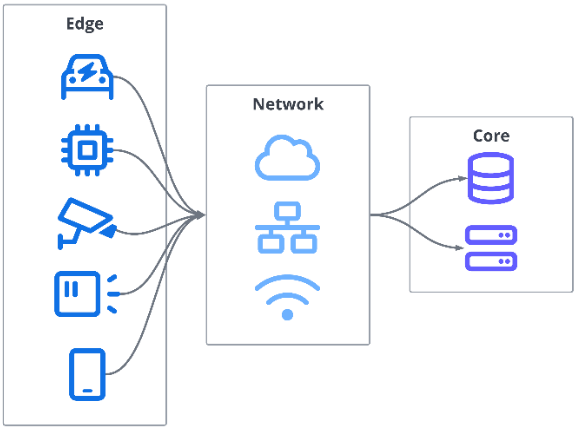

Figure 1. Data flows from edge to core

These trends indicate a significant need for more computing resources in public and private cloud datacentres, i.e., centralised computing resources. Transferring, storing, and processing data from edge devices centrally requires an increase in centralised computing resources proportional to the growth in edge computing resources and the growth of edge computing data utilisation. (Seagate, 2020)

Therefore, the main determinants of processing power required at core (i.e., data centres) following this conceptualisation are:

1) the data produced by edge devices, and
2) the percentage of data transferred from edge to core.

The percentage of data transferred from edge to core is used as the moderating variable for the relationship between data created on the edge (independent variable) and data centre spending (dependent variable).

Data transfers in the other direction, from core to edge, do not need to be analysed separately for the purpose of this research. Formulating a response and sending it over the network infrastructure to edge devices can be considered a part of the overall computing activity of the centralised computing resources in response to data generated by the edge devices. As such, data transfers in the direction of edge devices are mostly relevant when analysing the effects on network infrastructure, which is outside the scope of this research.

3.2 Causal model between edge computing and centralised computing

Growth of edge computing can be linked to more data being produced and consumed by the devices used for edge computing, therefore having a complementary effect on the spending on centralised computing. Since the consumers and producers of data are not necessarily collocated, processing needs to happen in network interchanges and data centres in order to connect data producers to data consumers (Buck Consultants International, 2021).

The causal link between data produced at the edge and the increase of processing power at core stems from this need to process, aggregate, and store the incoming data from edge in order to increase its practical value in many computing scenarios and use cases (Bilal et al., 2018). To minimise the effect of confounding variables, such as market growth in core computing, the model developed in this research focuses exclusively on core computing resources used to process data created and transmitted from the edge. As such, there is a direct link between the data created and transferred from the edge and core computing resources used to process this data.

The model is based on the assumption that there exists a 1:1 correlation between data transferred from edge to core and the data volume from edge that is processed at core. While there are scenarios where data may be processed multiple times, for example after being stored for a longer period of time as is the case in certain computing scenarios (Bilal et al., 2018), this is in most cases no longer relevant for the model, as such data may be regarded as not falling under the definition of data created and transferred from the edge. Due to the nature of digital data, it is impractical to distinguish between data originating from the edge and other data stored at core. The model therefore focuses only on the primary processing of data as it is transferred from edge to core.

The direct result of core computing resources being needed to process the incoming data from the edge is increased spending on core computing resources. Processing data in data centres is associated with various costs including electricity, hardware, software, personnel, and real estate for the data centre. This is reflected in the historical positive correlation between the data centre spending and the total volume of data processed in data centres as seen on Figure 2 on page 5 (Statista, 2021; Synergy Research Group, 2022). This research therefore links the increased use of core computing resources for the purposes of processing edge computing data to an increase in spending on core computing resources.

It is important to consider that adoption of edge computing as a paradigm can also mean that some computation no longer happens using centralised computing resources such as data centres, and data is instead processed and analysed close to where it is produced. Parts of data, either in raw of transformed form (as processed or aggregated analytics, etc.) will still need to be processed in centralised locations, for example because the data from an edge computing device needs to be combined or aggregated with other data in order to be useful or because it needs to be stored in the long-term for future use, or due to lack of feasible options for processing the data on the edge due to large volumes or complexity. However, much of the data produced on the edge will be processed on the edge directly, without needing to be collected and processed centrally. This has conceptually a negative effect on the demand for centralised computing, as this computing demand would be replaced with demand for edge computing resources (substitution effect).

3.3 Spending on centralised computing for processing edge data

While total spending on data processing has increased considerably year-on-year (Figure 2), processing a unit of data has become cheaper thanks to technological advances and economies of scale. $72.9 billion was spent on core computing in 2010 to process an estimated 2 ZB of data worldwide, while $218 billion was spent in 2020 to process an estimated 64.2 ZB of data (Synergy Research Group, 2022; Statista, 2021). This represents a reduction in spending per ZB of data processed from $36 bn / ZB in 2010 to only $3.4 bn / ZB in 2020 (Table 1). However, the decrease in spending per unit of data processed has slowed down in recent years as the gains in efficiency and technological advances produce diminishing returns (Figure 3). According to industry experts and previous research, it is likely that this trend will continue and the gains in efficiency will continue to decrease further (Koot & Wijnhoven, 2021).

Using a regression analysis, we find that the data centre spending per unit of data processed (y) follows the equation

$y = 34.836x^{-0.918}$ ($R^2 = 0.9908$),

with x representing the years since 2010, which is the year of the earliest available data (Figure 2). This formula can be used to forecast future spending per unit of data processed in order to establish a link between the volume of data created on the edge and associated spending on core computing to process that data.

Table 1. Spending on centralised data processing worldwide

Year	Data processed (ZB) (Statista, 2021)	Data centre spending (bn $) (Synergy Research Group, 2022)	Spending / data processed (bn $ / ZB)
2010	2.0	72.9	36.45
2011	5.0	81.0	16.20
2012	6.5	87.5	13.46
2013	9.0	91.0	10.11
2014	12.5	97.0	7.76
2015	15.5	103.0	6.65
2016	18.0	110.5	6.14
2017	26.0	128.0	4.92
2018	33.0	162.0	4.91
2019	41.0	189.3	4.62
2020	64.2	218.5	3.40

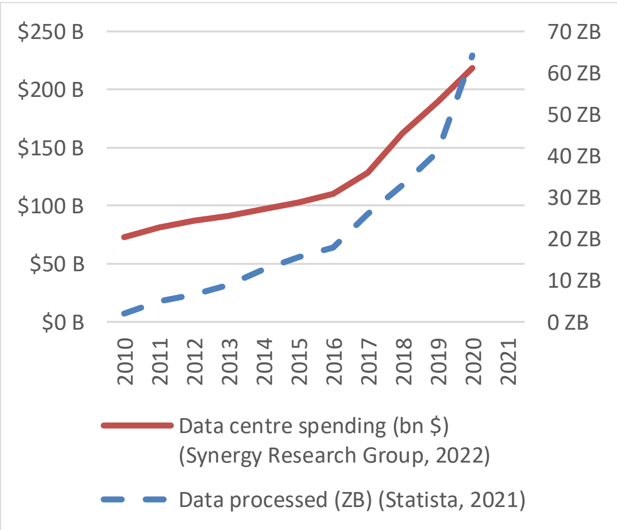

Figure 2. Spending on data centres and total volume of processed data

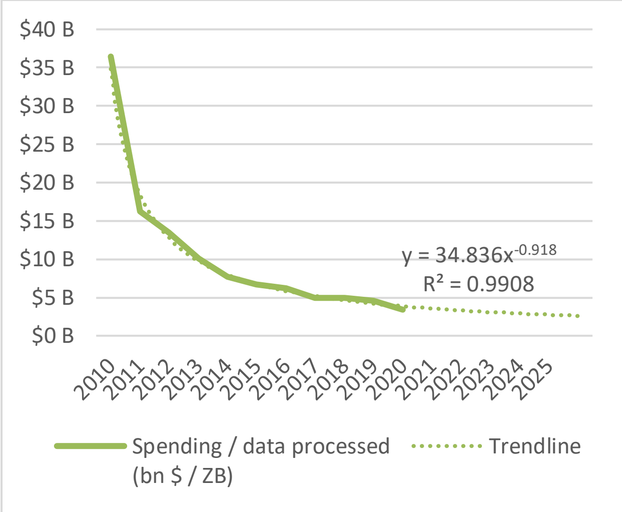

Figure 3. Worldwide data centre spending per processed ZB of data

This positive correlation between data centre (i.e., core) spending and the amount of data processed allows us to analyse the costs associated with processing data created at edge which is transferred to core. The core spending per processed data unit can be used to forecast the estimated additional spending on data centres resulting from the data created at edge. This assumes that the historical trend continues at least for the duration of the timeframe of this research, which is not unreasonable considering the relatively short timeframe of the forecast (5 years).

4. DEVELOPING THE MODEL

4.1 Variables

To develop the forecasting model, the following variables are used (Table 2):

- Data created on the edge per year (in zettabytes – ZB) (independent variable for edge computing prevalence)
- Data centre spending on processing edge data (in billions of USD) (dependent variable)
- Data transfers from edge to core (in % of data created) (moderating variable)
- Spending per unit of data processed (in billions of USD per ZB) (moderating variable)

To forecast future values, the yearly growth (or decline) rate of the independent variables must be included in the model:

- Change in data transfers from edge to core (in % per year)
- Change in data created on the edge per year (in % per year)

Table 2. Forecasting model variables

Variable	Unit	Source
Data created at edge per year	Zettabytes (ZB)	Seagate, 2020
Data centre spending on processing edge data	Billions of USD (bn $)	*Simulation*
Data transfers from edge to core	% of data created at edge	Seagate, 2020
Spending per unit of data processed	Billions of USD per ZB (bn $ / ZB)	Statista, 2021; Synergy Research Group, 2022
Change in data transfers from edge to core	% per year	Seagate, 2020
Change in data created on the edge	% per year	Seagate, 2020

The relationships between these variables can be visualised as seen on Figure 4.

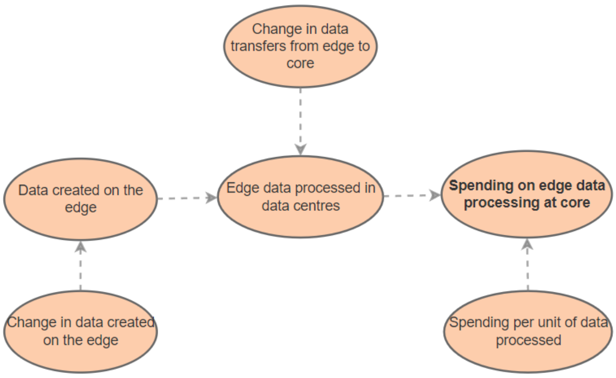

Figure 4. Data centre spending causal model

The task of core computing resources (i.e., data centres) in this context is to process the data created on the edge that is transferred to the core. As such, the data volume that is created can be used as the main input for quantifying the prevalence of edge computing for the purposes of this research. The volume of data created in 2020, which is the starting point of the simulation, was 18.2 ZB (Seagate, 2020).

Other ways to quantify edge computing prevalence were considered but were discarded for various reasons. The main alternative to using the created data volume is to use edge computing spending, similarly to how centralised computing demand is quantified. However, while there is a reasonable link between computing power and spending in the context of centralised computing, as data centres do not usually fulfil secondary tasks, the relationship between edge computing spending and computational resource demands (and therefore its effect on centralised computing demand) is not as simple. Edge computing is often used to enhance or augment a product which fulfils multiple secondary tasks. For example, a car costing €70 000 which includes an on-board chip to provide self-driving capabilities fits perfectly into the edge computing paradigm. However, it would be highly misleading to count this as €70 000 spent on edge computing, as the value of the car is made up of many different factors, and it is impossible to clearly separate the value of the functionality provided by the edge computing device (the self-driving capability) from the overall product. In other cases, the value provided by edge computing could be the dominant factor in the total spending amount. More importantly, a more expensive product with functionality provided by edge computing devices does not necessarily translate to higher compute requirements. As such, using edge computing spending would lead to highly misleading and non-actionable results.

The moderating variable, i.e., the percentage of data transferred from edge to core, is important to realistically model the effect that data created on the edge has on demand for core computing resources. Data created but not transferred to core does not have an effect on the demand for core computing resources. Therefore, the lower percentage of data transferred from edge to core, the smaller the effect that data created on the edge has on data centre spending. In 2020, 36% of data created at edge was estimated to be transferred to core for processing (Seagate, 2020). As it is unlikely for all of the data generated at the edge to ever be useful or practically viable for centralised analysis, the cap for this value is set at 80%.

The spending per unit of data processed follows the regression model based on developments between 2010 – 2020 (Figure 3):

$y = 34.836x^{-0.918}$

The model aims to forecast data centre spending resulting from the need to centrally process data created on edge. This is achieved by linking the volume of data transferred from edge to core for processing (Seagate, 2020) with forecasted spending per unit of data processed at core (Statista, 2021; Synergy Research Group, 2022). (Figure 4)

The simulation model is constructed using the online tool Insight Maker. The full list of equations and parameters as exported from Insight Maker and with the values based on scenario 1 (section 5.1) is shown in Figure 5. Values in other scenarios are changed as indicated in the beginning of each scenario.

Simulation Settings	**Time Start:** 0 **Time Length:** 5 **Time Step:** 1 **Time Units:** Years **Algorithm:** RK1

Model Variables	
Change in data created on the edge	**Value:** 42.2 **Units:** Unitless
Change in data transfers from edge to core	**Value:** 15 **Units:** Unitless
Data created on the edge	**Value:** 18.2 * (1 + [Change in data created on the edge]/100) ^ Years() **Units:** Unitless
Edge data processed in data centres	**Value:** (Min(35 * (1 + [Change in data transfers from edge to core]/100)^Years(), 80)/100) * [Data created on the edge] **Units:** Unitless
Spending on edge data processing at core	**Value:** [Edge data processed in data centres] * [Spending per unit of data processed] **Units:** Unitless
Spending per unit of data processed	**Value:** 34.836*((Years()+9)^-0.918) **Units:** Unitless

Figure 5. Model equations and values based on scenario 1

4.2 Timeframe

The model will explore developments in the 5 years after the initial objective data was collected. This time span was chosen due to the rapid pace of changes and development in the industry, which makes longer term forecasts inherently prone to uncertainty and inaccuracy. Due to the unavailability of data for the year 2021 at the time of conducting this research, the model will use values from 2020, therefore the analysis period is 2020 - 2025.

5. SCENARIOS

Based on the developed model, several scenarios can be constructed, simulated, and analysed to provide insight into future demand for centralised computing based on the growth of edge computing and its usage patterns (transfer rate from edge to core). The scenarios are selected based on predictions by industry experts and potential relevance for decision-makers relying on this paper to plan future spending on core computing.

5.1 Rapid proportional increase in centralised computing demand – complementary effect

Scenario assumptions / inputs:

- Change in data transfers from edge to core: 15% / year (Seagate, 2020)
- Change in data created on the edge: 42% / year (Seagate, 2020)

Table 3. Scenario 1 results

Year	Data created on the edge	Edge data processed in data centres	Spending on edge data processing at core	Spending per unit of data processed
2020	18 ZB	6 ZB	$29.5 bn	$4.6 bn/ZB
2021	26 ZB	10 ZB	$43.8 bn	$4.2 bn/ZB
2022	37 ZB	17 ZB	$65.5 bn	$3.9 bn/ZB
2023	52 ZB	28 ZB	$98.7 bn	$3.6 bn/ZB
2024	74 ZB	45 ZB	$149.8 bn	$3.3 bn/ZB
2025	105 ZB	74 ZB	$228.5 bn	$3.1 bn/ZB

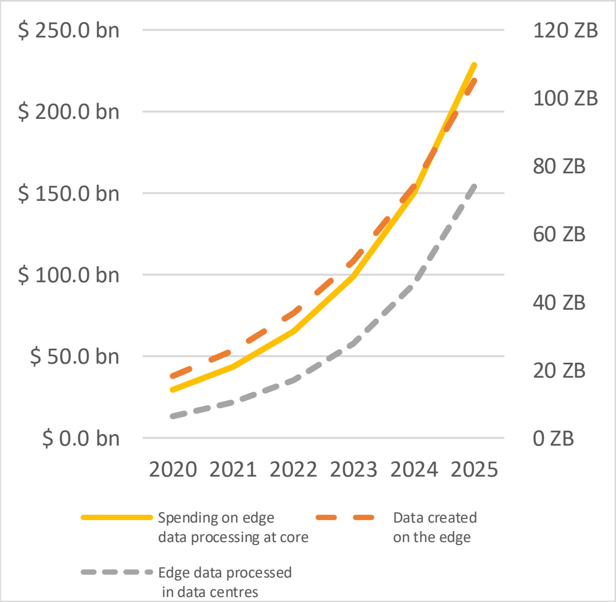

Figure 6. Scenario 1 results visualisation

In this scenario, the amount of data generated on the edge and the ratio of data transferred from edge to core grows rapidly, as currently forecasted by industry experts such as Seagate (2020). (Table 3)

Considering the need to transfer significant amounts of data from the edge to the core for processing and analysis, and the need for compute resources to perform the analysis, edge computing growth and changes in behaviour regarding enterprise data processing could lead to a significant increase in the growth of centralised computing demand. (Figure 4)

The growth of data transfers from edge to core is set as 15% per year, based on research by IDC as reported by Seagate (2020).

The growth of data created on the edge is set at 42% per year, based on forecasts in the report by Seagate (2020). This is a very fast pace of growth and would mean the volume of data increasing to nearly 105 ZB per year by 2025. (Table 3)

For this scenario to realise, the trend of transferring increasingly more data from the edge towards the core as described in section 3.1 continues. The increased data volumes and the need for aggregating, storing, and analysing the data from edge computing devices would necessitate an increase in centralised computing resources to meet this demand. Therefore, an increase in edge computing demand would lead to a proportional increase in centralised computing demand for the purposes of processing the data transferred from the edge, reaching $228.5bn by 2025.

This scenario would require planning for a significant increase in cloud or data centre spending by enterprises who rely on or plan to expand their edge computing resources, in order to keep up with the significant increase in data volumes and to be able to use the data generated by edge computing devices to their advantage most effectively.

5.2 Continued historical growth rate of data creation

Scenario assumptions / inputs:

- Change in data transfers from edge to core: 15% / year (Seagate, 2020)
- Change in data created on the edge: 23% (Statista, 2021)

Table 4. Scenario 2 results

Year	Data created on the edge	Edge data processed in data centres	Spending on edge data processing at core	Spending per unit of data processed
2020	18 ZB	6 ZB	$29.5 bn	$4.6 bn/ZB
2021	22 ZB	9 ZB	$37.9 bn	$4.2 bn/ZB
2022	28 ZB	13 ZB	$49.1 bn	$3.9 bn/ZB
2023	34 ZB	18 ZB	$64.2 bn	$3.6 bn/ZB
2024	42 ZB	26 ZB	$84.3 bn	$3.3 bn/ZB
2025	51 ZB	36 ZB	$111.4 bn	$3.1 bn/ZB

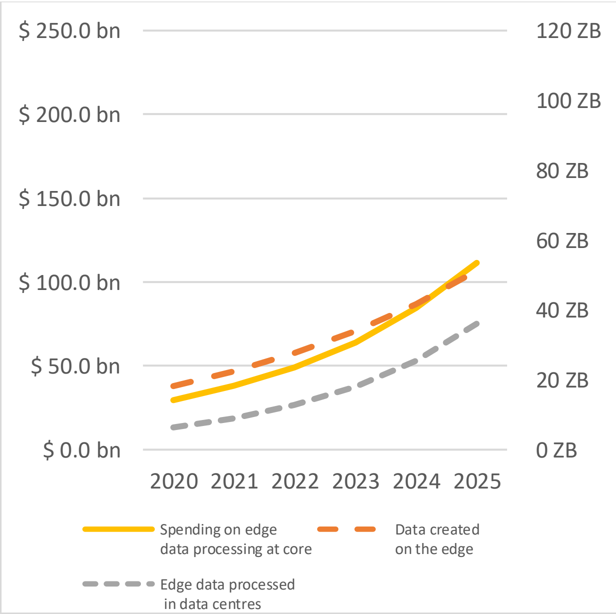

Figure 7. Scenario 2 results visualisation

This scenario explores the possibility that the current very ambitious growth forecasts for edge computing do not realise to such an extent as expected, and instead follows the same trend as historically shown by the total volume of data processed at core. At the same time, data transfers from edge to core increase as forecasted by experts. As such, the scenario assumes that data transfers from edge to core increase by 14% per year (Seagate, 2020) and data created on the edge increases by a CAGR of 23%.

In this scenario, data transfers from edge to core increase at a slow pace as they have thus far, as consumers struggle to effectively aggregate and make use of the data generated on the edge. The result is a much more modest increase in both data created on the edge, which would reach 51 ZB by 2025, as well as data that is processed at core, at 36 ZB by 2025.

The modest increase in data transfers combined with the 10% annual increase in created data on edge would still have a noticeable effect on the demand for centralised computing resources but would not mean an exponential increase as seen in Scenario 1. (Figure 7)

The result of this scenario is data centre spending on processing edge data increasing to $111.4 billion USD by 2025 (Table 4). This is a significant growth, and companies planning to invest in edge computing and expecting the historical growth rate would need to take these additional expenses into account.

5.3 Edge computing partially substitutes centralised computing

Scenario assumptions / inputs:

- Change in data transfers from edge to core: -10% / year
- Change in data created on the edge: 42% / year (Seagate, 2020)

Table 5. Scenario 3 results

Year	Data created on the edge	Edge data processed in data centres	Spending on edge data processing at core	Spending per unit of data processed
2020	18 ZB	6 ZB	$29.5 bn	$4.6 bn/ZB
2021	26 ZB	8 ZB	$34.3 bn	$4.2 bn/ZB
2022	37 ZB	10 ZB	$40.1 bn	$3.9 bn/ZB
2023	52 ZB	13 ZB	$47.3 bn	$3.6 bn/ZB
2024	74 ZB	17 ZB	$56.2 bn	$3.3 bn/ZB
2025	105 ZB	22 ZB	$67.1 bn	$3.1 bn/ZB

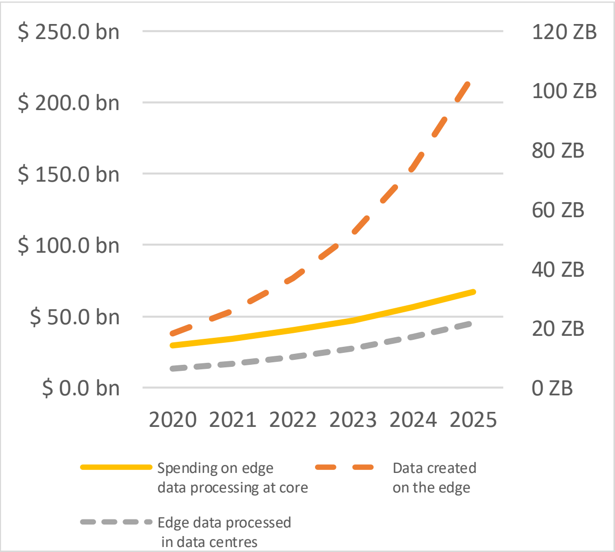

Figure 8. Scenario 3 results visualisation

The shift of investments and industry focus from centralised computing to edge computing could be significant enough to have a negative effect on the demand of centralised computing. In this scenario, enterprises would shift significant compute workloads away from the core and towards the edge, which renders some centralised computing resources obsolete and unnecessary. If this shift happens relatively simultaneously across the industry, it could lead to a decrease in investments into centralised computing, causing the demand to drop.

This scenario assumes that the trend of increased data transfers from the edge to the core (section 2.7) does not realise to a significant extent, and that the shift of centralised computing workloads towards the edge is significant enough to counteract the need for more centralised computing resources for processing and analysing data from the edge. This leads to the assumption that data transfers from edge to core will decrease by 10% per year as there is less need for processing data centrally. This value is based on a hypothetical scenario and is not supported by historical data.

In this scenario, spending on centralised computing would begin to decline as resources are shifted to edge computing instead. This is exacerbated by the ratio of data transferred from edge to core also declining as more processing is done directly on the edge with less central aggregation. As a result, centralised computing spending on processing edge data would only increase to $67.1 billion USD by 2025 (Table 5), far below the values forecasted in scenarios 1 and 2.

Due to the increased use of edge computing, data created per year at the edge increases at a rapid pace in the same way as in the first scenario (section 4.1). However, since only a small portion of this data is transferred to the core, this does not have a sufficient complementary effect on the spending on centralised computing resources. (Figure 8)

5.4 Discussion of the scenarios

Each of the three scenarios explores a potential future regarding the amount of spending on centralised computing resources for processing edge data. They all forecast some increase in this spending, even in a pessimistic scenario where centralised computing is partially substituted by edge computing. As such, it is vital for companies that invest in edge computing or are involved with the processing of edge data to consider the associated costs related to processing this data in their spending plans.

The first scenario (section 5.1) examines the possibility that both the volume of data created on the edge and the rate of data transfers from edge to core continue to grow at a rapid pace. This is the scenario predicted by Seagate and IDC (2020) and is extremely favourable towards edge computing and the need to centrally process data produced at edge. The likelihood of this scenario realising may be further increased by the effects of the COVID-19 pandemic, which could have fuelled a global increase in the digitalisation of communication, thereby potentially drastically increasing the amount of data generated and transferred on the edge. However, the full effects of the pandemic on the prevalence of edge computing are not yet fully known and may present an opportunity for further research.

Scenario 1 is an interesting scenario for companies that see the benefits of edge computing as potentially beneficial for their operations and plan to invest in edge computing more heavily in the future, as the global scenario will likely mimic the local changes in centralised computing demand in their own company. It shows that despite the decrease of per-unit data processing costs at core, the volume of data created at edge is potentially significant enough to cause large increases in spending which must be accounted for.

The second scenario (section 5.2) assumes that the growth of data creation on the edge follows the historical trend of growth of total data created and processed in the 5 years preceding the analysis period, from 2015 – 2020 (Figure 2), which means a compound annual growth rate (CAGR) of 23% (Statista, 2021). This means that the scenario will explore a less drastic change in data creation on edge based on historical data, while assuming the same rate of increase in data transferred from edge to core based on expert forecasts. However, industry trends and changes in use cases may lead to a different growth pattern in edge computing compared to the overall data volume. The strong industry focus on developing new solutions and applications of edge computing (Bilal et al., 2018), and the large variety of potential use cases may challenge the historical growth rates. Whether the historical assumption holds true or not, it is important for companies to acknowledge the significant rise in core computing demand even in this more conservative scenario.

Despite the growth of edge computing data creation following historical trends in this scenario, it is important for companies to consider the need for core computing resources beyond what they are expecting based on workloads which do not include edge computing data. The increase in data transferred from edge to core may pose a significant increase in the total workload, as it would complement the existing demand for core computing resources.

The third scenario (section 5.3) analyses the potential for a substitution effect, in which edge computing continues to grow, but data transfers to core decrease as computing resources are shifted from core to edge. This scenario is not the most likely, as the prevailing consensus of industry experts is that core computing will continue to support edge computing and both paradigms will coexist in the future. However, based on the multitude of use cases of edge computing, it is not impossible that at least a partial substitution could occur if external factors cause edge computing to become cheaper or new computing use cases rely heavily on the benefits provided by edge computing, such as low latency and proximity to user (Ren et al., 2018). This scenario is most relevant for those who plan to exclusively invest in or service edge computing applications which do not benefit from data aggregation and centralised data processing.

However, it is remarkable that even in a scenario where data transfers decrease, the third scenario still forecasts a significant increase in demand and therefore spending on core computing resources. It is therefore vital that a significant growth in edge computing prevalence is not mistakenly associated with core computing becoming unnecessary or obsolete, even if proportionally less data is transferred to the core. Even under a pessimistic scenario, core computing remains vital as the means of aggerating and processing edge data. This implies that further investments in core computing remain important despite the growing prevalence of edge computing.

6. CONCLUSION

The increased prevalence and proliferation of edge computing as a paradigm poses both opportunities and challenges for traditional IT infrastructure and cloud data centres. It is impossible to predict the exact effects of these changes in how data is collected, consumed, and used, but enterprises can benefit from taking the potential future scenarios into account when planning their spending and activities.

The research model constructed in this research incorporates the most relevant indicators for the growth of edge computing prevalence, centralised computing demand, and their mutual effects to enable constructing potential scenarios to aid in decision-making and forecasting.

This research identified three possible future scenarios for the effects of edge computing growth on the demand for centralised computing resources which help enterprises make the right decisions when formulating their IT strategy.

The research further helps the understanding of the effect between edge computing growth and its effect on the demand of core computing resources to process the data created on the edge.

The first and most likely scenario is based on the assumption that the adoption of edge computing continues rapidly over the next 5 years, with 105 ZB of data created annually at edge, and that a large part of this data needs to be transferred to centralised computing resources for processing. This leads to $228.5bn of expected spending on centralised computing resources for processing edge data by 2025, necessitating significant investments by enterprises utilising edge computing.

The second scenario is based on the assumption that edge computing adoption follows the historical trends of data creation, and data produced at the edge grows at a slower pace than currently anticipated by industry experts, increasing to 51 ZB annually in the next 5 years. This leads to a correspondingly slower increase in centralised computing demand, with the spending reaching $111.4bn by 2025, but no substitution effect occurs, and the centralised computing market for processing edge data continues to grow.

The third scenario analyses a potential substitution effect, whereby edge computing continues to grow rapidly while becoming more isolated from centralised computing, taking over some of its tasks. As such, data transfers from edge to core decrease and demand for centralised computing for the purposes of processing edge data increases at a much lower pace as investments shift exclusively towards edge computing instead, with the forecasted spending at $67.1bn by 2025. However, full substitution does not occur, and centralised computing demand continues to grow compared to 2020 in order to process the large volumes of edge data.

This research provides practical implications for managers and investment-related decisionmakers, with the key takeaway being the continued growth of demand for centralised computing resources in every projected scenario, even if edge computing starts to substitute core computing in some areas.

Future research in this area could focus on the implications on specific resources needed for the processing of edge data, and the balance between processing and storage of the data. While this research takes a spending-focused approach, research focusing on specific computing resource use could provide more useful data for server hardware manufacturers and data centre operators.

Additionally, research could be broadened on the differences in edge computing usage per industry and how these differences affect the demand for core computing resources for edge data originating from these industries. It is likely that different industries have different needs and usage patterns for edge computing, resulting in varying demand for core computing resources.

7. REFERENCES

- Bilal, K., Khalid, O., Erbad, A., & Khan, S. U. (2018). Potentials, trends, and prospects in edge technologies: Fog, cloudlet, mobile edge, and micro data centers. Computer Networks, 130, 94–120. https://doi.org/10.1016/j.comnet.2017.10.002
- Buck Consultants International. (2021, June). Verkenning relatie accommoderen datacentervraag en digitaliseringskansen. Dutch Ministry of Economic and Climate Affairs. https://www.rijksoverheid.nl/documenten/rapporten/2021/06/14/verkenning-relatie-accommoderen-datacentervraag-en-digitaliseringskansen#:~:text=hier%3A%20Home%20Documenten-,Verkenning%20relatie%20accommoderen%20datacentervraag%20en%20digitaliseringskansen,uitgevoerd%20door%20BUCK%20Consultants%20International
- Gartner. (August 2, 2021). Public cloud services end-user spending worldwide from 2017 to 2022 (in billion U.S. dollars) [Graph]. In Statista. Retrieved April 18, 2022, from https://www-statista-com.ezproxy2.utwente.nl/statistics/273818/global-revenue-generated-with-cloud-computing-since-2009/
- GSMA . (July 7, 2021). Market share of mobile telecommunication technologies worldwide from 2016 to 2025, by generation [Graph]. In Statista. Retrieved April 18, 2022, from https://www-statista-com.ezproxy2.utwente.nl/statistics/740442/worldwide-share-of-mobile-telecommunication-technology/
- Statista. (2021, June). Amount of data created, consumed, and stored 2010–2025. Retrieved June 27, 2022, from https://www.statista.com/statistics/871513/worldwide-data-created/
- IDC, & Statista. (January 13, 2022). Edge computing market revenue worldwide from 2019 to 2025 (in billion U.S. dollars) [Graph]. In Statista. Retrieved April 18, 2022, from https://www-statista-com.ezproxy2.utwente.nl/statistics/1175706/worldwide-edge-computing-market-revenue/
- Khan, W. Z., Ahmed, E., Hakak, S., Yaqoob, I., & Ahmed, A. (2019). Edge computing: A survey. Future Generation Computer Systems, 97, 219–235. https://doi.org/10.1016/j.future.2019.02.050
- Koot, M., & Wijnhoven, F. (2021). Usage impact on data center electricity needs: A system dynamic forecasting model. Applied Energy, 291. https://doi.org/10.1016/j.apenergy.2021.116798
- Nguyen, T. A., Min, D., Choi, E., & Tran, T. D. (2019). Reliability and Availability Evaluation for Cloud Data Center Networks Using Hierarchical Models. IEEE Access, 7, 9273–9313. https://doi.org/10.1109/access.2019.2891282
- Ren, J., Zhang, D., He, S., Zhang, Y., & Li, T. (2020). A Survey on End-Edge-Cloud Orchestrated Network Computing Paradigms. ACM Computing Surveys, 52(6), 1–36. https://doi.org/10.1145/3362031
- Seagate. (2020). Rethink Data Report 2020. https://www.seagate.com/files/www-content/our-story/rethink-data/files/Rethink_Data_Report_2020.pdf
- Synergy Research Group. (2022). Enterprise spending on cloud and data centers by segment from 2009 to 2021. https://www.srgresearch.com/articles/as-quarterly-cloud-spending-jumps-to-over-50b-microsoft-looms-larger-in-amazons-rear-mirror