JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A MODEL TO DETERMINE THE SUSTAINABILITY OF CLOUD SERVICE PROVIDER

Authored By

Daksh Kanoria Rohith KandlaguntaSaket Gupta
Prof. P. J. Jambhulkar

DEPARTMENT OF COMPUTER ENGINEERING
Pune Institute of Computer TechnologyDhankawadi, Pune - 411043

ABSTRACT

There has been a double-digit growth in the number of devices that are connected to the internet. These devices range from autonomous cars to im- plants in the Brain. It is safe to say that a lot of data is being generated by these devices. This has led to a dramatic increase in the resources needed to build, run and maintain these data centers. There is no denying that we are better off with all the new compute capacity that has been created. However, will this capacity increase come at the price of our environment? We must ensure that new industries as well as old industries are being transitioned to sustainable energy, water and other natural resources.

Using various different metrics, we have identified 4 different metrics that affect the sustainability of a Cloud Service Provider. We have the one of them to create buckets based on the size of the data centres and used the other 3 to create a metric that can calculate their sustainability.

CHAPTER 1 INTRODUCTION

1.1 Overview

"Data is the new Oil". Data and information have become an integral part of our lives. For most of the world's population, how and where the data has been stored, is not even the question that crosses their mind. Users are more concerned with the quality of service rather than the infrastructure behind it.

During last few years there has been a dramatic increase in the number of devices that generate data. Internet traffic has grown exponentially in the past few decades from just a couple of TBs in 1987 to 60 PB in 1997, 50 EB's in 2007, and finally 1.1 ZB's in 2017. This increase in internet traffic and the accompanying data that need to be stored has led to a dramatic increase in the number of data centers and cloud infrastructure providers. Amazon, Google, Microsoft, and other major IT service providers have at least double-digit growth in the number of servers that they have deployedto manage this level of generated data.

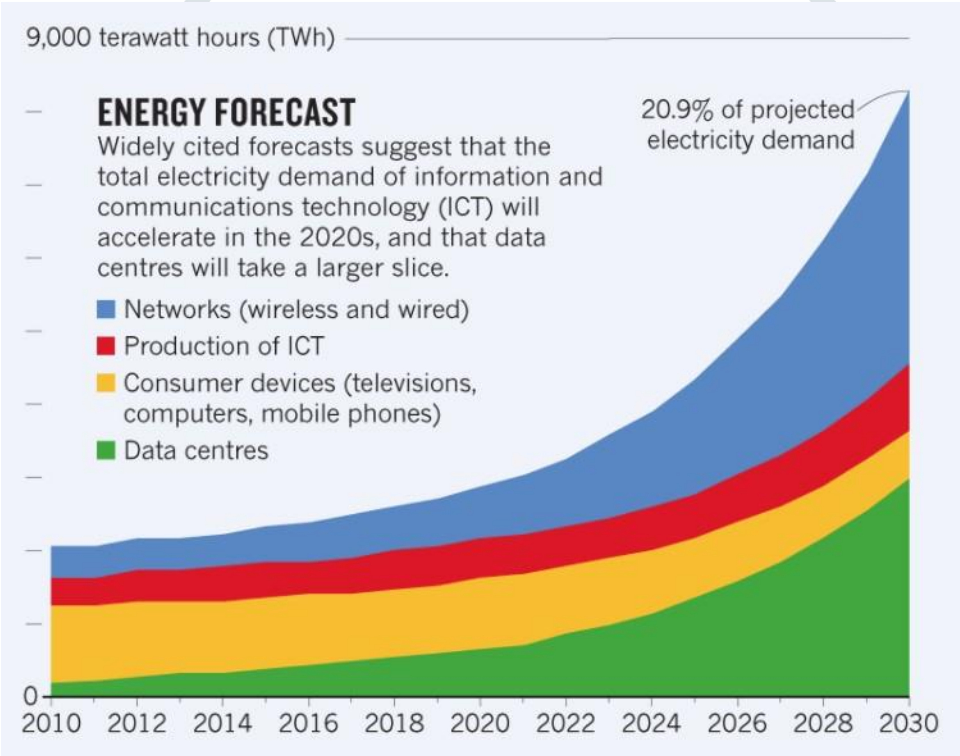

Figure 1.1: Growth in Energy Usage[1]

In 2018, it was estimated that the global data center energy usage rose to 205 TWs. or around 1 percent of global electricity consumption. This represents a 6 percent increase compared with 2010, whereas global data center computes instances increased by 550 percent over the same period. Even though there has been a 4x increase in the efficiency of the processors, a considerable number of resources and energy has been utilized in cooling and powering these network devices such as servers, storage, and networkinginfrastructure.

All the benefits and new services that these cloud data center have en- abled must not come at the price of the environment. We must ensure that the growth in the cloud infrastructure is being done sustainably. Natural resources such as water, air, coal, and natural gas are rapidly being utilized to power these data centers. All the major cloud service providers like AWS, Microsoft Azure, etc. have been switching to renewable resources such as solar or wind energy for electricity.

Sustainability in the data center is not just about where the power is coming from. It also includes the water that is being to cool the data center and its server's, carbon and other harmful gases that are being emitted dur- ing the operation and recycling of components like servers after they have reached, they end of life. There is an endless list of metrics that can be looked at the determine the sustainability of a data center.

With business and consumers becoming more aware of their environ- mental impact. They should also be able to see the impact that the servers and storage they are using from these cloud service providers in having on the environment. They should also be allowed to choose and define the met- rics that are important to them. Some may prefer CPS's that use less water. Some may prefer CPS's that emit less carbon dioxide.

In this study we aim to investigate the following questions: (i) What are the factors have a taggable impact on the environment during the opera- tions of a CSP? (ii) Is it possible to create a simple metric that can be usedto measure the suitability of CSP?

1.2 Motivation

New technologies like augmented reality, work from home, social media are all increasing our dependence on technology. A common thread between all of the new technologies that are quickly becoming essential; is the need to be connected to the internet. The internet is nothing but a network of interconnected computers.

At the same time sustainability has become an important part of any person's life. Knowingly or unknowing each interaction we have has an im- pact on the environment. There has been a bit of push by individuals and organizations to make more sustainable directions. These individuals also want to know about the steps that are being taken by the companies whose products and services they use; in terms of the environmental impact these companies have.

However, everyone can't do the research and find an answer that sat- isfies them. Like any complex system, sustainability has a lot of different factors to be taken into consideration, for example, location, time of year, number of users, source of energy, up-time, etc. To make an informed choice means doing a lot of research on these factors and trying to understand them.

To make this process easier and to introduce some transparency into this process we wanted to come up with a model that would take into account several different factors, weigh them according to the impact they have, and come with a definite score that would let any individual or organization knowhow sustainable their cloud infrastructure provider is.

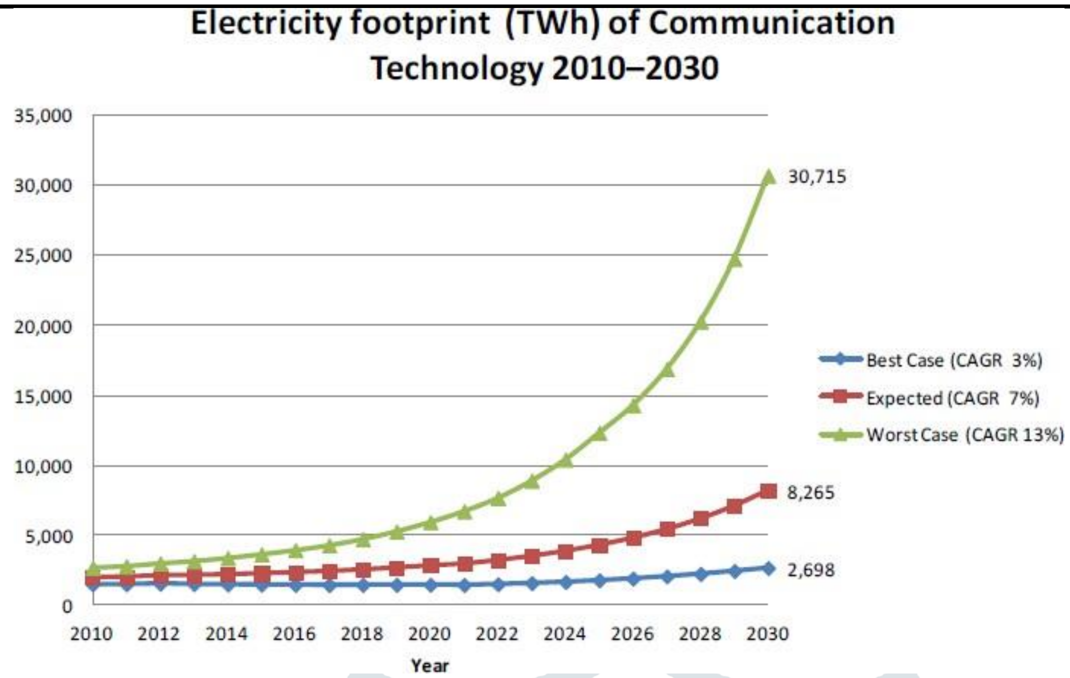

Figure 1.2: Global electricity demand of communication technology2010–2030[1]

CHAPTER 2 METHODOLOGY

2.1 Overview

We will be looking at both the direct and indirect factors that have an impact. Our goal is to identify these factors and find a suitable way of ranking the impact each individual factor has on the environment.

2.2 Dataset

In an ideal sicario we would like to collect as much data as possible. Metric of Sustainable Data Center have been discussed by Reddy et al (add reference). As there is no complete dataset that account for both the direct and indirect impact a data center used by a CSP. We collected as much in- formation publicly available. Factor for which we could not find reliable data sources we have generated that information with reference to the work done by Siddik el al[6].

2.3 Identification of factors

There is a long list of factors both direct and indirect that can be used to determine how sustainable a CSP is. We choose to use a mix of direct and indirect factors. Since water and energy are required in very large quantities.

Water is mainly used for cooling purposes. Data Centers in arid region cannot consume the same amount of was as Data Centers in region with high availability of water. There is also and energy cost associated with the treatment and transportation of water to these Data

Centers. Water from scarcity sources as defined by ISO 14046 have a much greater impact on theirlocal region.

Energy is the other main resource needed for day-to-day operations. Energy can be from renewable and non-renewable sources. The impact re- newable sources have during their construction and operation are negative compared to their long-term impact compared to non-renewable sources. Hence, those will not be considered.

While data center themselves don't produce a lot of green house gasses. The sources that supply the water and energy to them produce a lot of green house gases. Especially CO2, which has been proven to have a taggable im-pact on the local and global climate.

PUE or is a key metric that a lot of CSP to talk about their sustain- ability. However, PUE was not designed to be used as a sustainability metric. However, PUE can tell us a lot about the efficacy of the equipment used by CSP.

There are also other factors both direct and indirect that we would have liked to include in our dataset. There is not enough data available pub- licly to be added to our data. However, with approach taken in this study these can be added to the dataset for analysis.

2.4 Factor Analysis

Principle Component Analysis is a common technique for evaluation of sustainability indicators. PCA takes into account the variation in the data. It also helps us in identifying sets of factors that have an impact. We found that using 3 Principal Component accounts for 80 percent of the variation in the data. After which we conducted Kaiser-Meyer-Olkin (KMO) test to measure the adequacy of the samples. A value of 0.72 was recorded. Based on the literature of the KMO test 0.72 is enough for use to processed.

Three components were generated which accounted for 0.547, 0.157 and 0.133 variance in the data respectively. As the rest of components did not match the threshold of 0.5, those where not considered for this study. We then identified the factors which met the loading threshold value.

2.5 Data Normalization

The factors that were identified were all in different magnitudes. We normalized these values using Standard Scaler function available in the Sklearnlibrary.

2.6 Metric Calculation

Based on the results of the various techniques and tests so far, we determined that it possible to create a single metric. We calculated our Sus- tainability Metric (SM) using

CHAPTER 3

RESULTS

3.1 Factors that affect sustainability of CSP

Based on the results of the various techniques and tests so far, we determined that it possible to create a single metric. We calculated our Sus- tainability Metric (SM) using

1. Gross Sqft – The size of the data center that is used by the CSP toprovided services
2. Water Usage (m3 MWh-1) – The water used by the data center forcooling and other purposes
3. Carbon Released (tons/MWh) – The amount of carbon released by thedata center and the other services needed for smooth operations.
4. Renewable Energy (percentage) – The amount of energy acquired fromrenewable sources such as solar or wind.

3.2 Weight of the factors on the metric

Factor (ii) to (iv) have a direct correlation to the suitability of the CSP. Taking factor (i) in to account in the Sustainability Metric will change the unit of the final metric. Hence, it will be omitted from the final calculation.

3.3 Importance of size and location of datacenter

We have observed that location of the data center has no direct co- relation with sustainability. Although, there are a few factors with indirectly influence it. If a data center is located where there is an abundance and access to cheap of renewable, will tent to use them, as compared to a data centers. Same thing can be said for size of the data centers.

Small Data centers require less resources to set up and run comp com- pared larger datacenters. Larger data center can be more efficient in term of resources utilisation and PUE.

3.4 Calculating Sustainability

Our final sustainability output for each data center will lie between negative infinity and positive infinity. This scale will be used to compare sustainability of different data centers. Ideally a data center should have sustainability of negative infinity. This means that it uses 100% renewable energy, no water and there is no waste emission. On the other hand a data center with sustainability as positive infinity will be using 0% renewable en- ergy, infinite water resources and infinite waste generation.

CHAPTER 4 CONCLUSION

After performing analysis on the data sets and after looking at thee results obtained, it can be concluded that there are certain factors such as percentage of renewable energy utilizes, quantity of water used and waste generated(via air), are the three major factors which have the most affect on sustainability.

It was observed that sustainability of a CSP is directly proportional to percentage of renewable energy utilize, inversely proportional to water usage and inversely proportional to waste emission. This was achieved using factoranalysis.

After obtaining the factors and performing appropriate scaling, equal weights were assigned to all the factors. Using the formula mentioned above, all the three factors were multiplied, and the result was considered as sus- tainability for that particular CSP.

CHAPTER 5

REFERENCES

[1] Masanet, Eric & Shehabi, Arman & Lei, Nuoa & Smith, Sarah & Koomey, Jonathan. (2020). Recalibrating global data center energy-use estimates. Science. 367. 984-986. 10.1126/science.aba3758.

[2] J. Roque, L. Chauvel, M. Aloqaily and B. Kantarci, ”A Feasibility Study on Sustainability-Driven Infrastructure Management in Cloud Data Cen- ters,” 2018 IEEE Canadian Conference on Electrical Computer Engi- neering (CCECE), 2018, pp. 1-4, doi: 10.1109/CCECE.2018.8447754.

[3] Md Abu Bakar Siddik et al 2021 Environ. Res. Lett. 16 064017

[4] Lai, Han, Huchang Liao, Jonas Šaparauskas, Audrius Banaitis, Fernando A.F. Ferreira, and Abdullah Al-Barakati. 2020. ”Sustainable Cloud Ser- vice Provider Development by a Z-Number-Based DNMA Method with Gini-Coefficient-Based Weight Determination” Sustainability 12, no. 8: 3410. https://doi.org/10.3390/su12083410

[5] Andrae, Anders S.G., and Tomas Edler. 2015. ”On Global Electricity Usage of Communication Technology: Trends to 2030” Challenges 6, no. 1: 117-157. https://doi.org/10.3390/challe6010117

[6] M. A. B. Siddik, A. Shehabiand L. Marston,“The environmental foot- print of data centers in the United States”.University Libraries, Virginia Tech,05-May-2021,doi:10.7294/14504913.v2.

[7] Boulay A-M et al 2018 The WULCA consensus characterization model for water scarcity footprints: assessing impacts of water consumption based on available water remaining (AWARE) Int. J. Life Cycle Assess. 23 368–78

[8] Mytton, David. (2021). Data centre water consumption. npj Clean Water. 4. 10.1038/s41545-021-00101-w.