Decision Support in Data Centers for Sustainability

Michael Pawlish[1,3] Aparna S. Varde[2,3] and Stefan Robila[2]
1. Department of Earth and Environmental Studies
2. Department of Computer Science
3. Environmental Management Program
Montclair State University, Montclair, NJ
(pawlishm1 | vardea | robilas)@montclair.edu

***Abstract*— In this paper, we propose a decision support system (DSS) for the greening of data centers to help the environment, hence promoting sustainability. As society continues the relentless shift towards electronic communications there is a growing demand for greater storage and processing on local and cloud data centers. This increasing demand on both storage and processing is calling for greater resource use in the form of electrical consumption and physical resources to build data centers. A potential area of improvement is to gain greater server utilization rates since traditionally the phenomenon of "server sprawl" occurs where more servers are added to the data center without seeking greater utilization rates on existing servers first. This implies maintaining more servers than actually needed that translates to greater carbon dioxide emissions causing potential environmental problems. Presently, average server utilization rates in most data centers are rather low, and we make the claim that utilization rates should be increased so that we can lower the number of servers for enhanced sustainability. This is the focus of our paper. We provide recommendations based on the DSS that would be useful to data center operators in academia and also in the industry.**

Keywords*— *Case-Based Reasoning, Data Centers, Decision Support Systems, Green Information Technology

I. I. INTRODUCTION

A growing environmental hazard today is climate change. Scientists around the world have identified that the earth is getting warmer resulting in a substantial change in weather patterns at the regional level. The change to a warmer climate has been sharply debated politically, but the vast majority of scientists have agreed that a cause is attributed to the increasing carbon dioxide levels in the atmosphere. In addition, increased carbon dioxide emissions into the atmosphere also cause environmental pollution that is another potential hazard. Carbon dioxide levels have been documented as increasing by scientists since the start of the Industrial Revolution due to the burning of fossil fuels. One rather recent contribution to the rising carbon dioxide levels is the data centers that make up the backbone for the Internet Revolution. While data centers provide an important component to communications, our paper will argue that data centers can be made more energy efficient through decreasing "server sprawl", a concept we explain next.

Server sprawl is a natural tendency by data center administrators to simply add more servers to the data center when new applications are developed rather than gaining higher server utilization rates by adding the new applications to existing servers. In addition, there is a current trend in data centers to have low server utilization rates that translates to the servers running, but not processing useful work. Our additional recommendation is to operate servers at a higher utilization rate to increase energy efficiency.

The process of mitigating environmental hazards comes down to the concept of seeking a sustainable solution. Sustainability science seeks real world solutions to sustainability issues and aims to break down artificial and outdated disciplinary gaps between the natural and social sciences through the creation of new knowledge and its practical application to technology transfer and decision-making [6, 12, 20].

The U.S. data center industry is in the midst of a major growth period stimulated by increasing demand for data processing and storage at all levels [19]. During the past five years, increasing demands for computer resources has led to significant growth in the number of data center servers, along with an estimated doubling in the energy used by these servers and the power and cooling infrastructure that supports them. Pursuit of energy efficiency opportunities in data centers is important because of the potential for rapid growth in direct energy use.

The proposed research is in the area of applying the concept of sustainability to data centers. It aims to enable better decision-making to create energy efficient data centers heading towards a greener and more sustainable environment. An important corresponding research goal is thus the development of a decision support system that will allow data center professionals to make better management decisions for server and other information technology systems while balancing energy efficiency with various functionality demands. While the focus is on data centers in academia, the outcomes of this research would also be useful in the corporate world where green energy is an important issue.

The rest of this paper is organized as follows. Section II defines the problem we address. Section III explains our proposed approach to solve the problem with details on data requirements and mining paradigms. Section IV outlines DSS questions and answers that depict prospective user queries and responses. Section V gives recommendations for users based on our analysis. Section VI overviews some related work in the area. Section VII states our conclusions and ongoing work.

II. II. Problem Definition

A. *Goals*

The main goal of this research is to conduct a detailed analysis of data centers from a sustainability energy perspective while also meeting the demands of accuracy and efficiency. An important related goal thereof is to develop a decision support system (DSS) for green computing in data centers. Decision support systems are used throughout the world today in various fields such as medicine, farming, management, routing airplanes, and others where complex organization is needed. More specifically, in this initiative, a DSS will be developed for the design and retrofitting of data centers.

B. Challenges

While energy efficiency is viewed as a desirable outcome, there is a risk aversion by system administrators to change that could possibly result in down time of data centers [19]. There also exists a bureaucracy problem in logistically gathering data between various departments in universities such as information technology (IT), facilities and accounting. This is due to the fact that in most universities, the IT operations provide data center services, the facilities departments provide the electricity to power the data center, and the accounting offices pay for the services. Therefore, a disconnect exists or a split incentive occurs, according to the Environmental Protection Agency (EPA), between the various departments since there is no single entity responsible for lowering electrical usage and carbon footprint reduction [18]. This poses a challenge with respect to bureaucracy and privacy issues, which is often encountered in data management and knowledge discovery.

Another challenge is the physical means for collection of real data. There is often a lack of devices such as sub-meters, in particular for the air-conditioning system. There is clearly a need to move from a mere estimation of air-conditioning costs and carbon footprint, to more exact measurements. To reach this goal, it will be necessary to install meters on the air-conditioning system in the near future, which presents a challenge in data collection.

III. Proposed Approach

We propose to build the required decision support system for green data centers by conducting analysis with data mining approaches such as case-based reasoning (CBR) and decision trees, while incorporating environmental management aspects such as thermal profiling and virtualization. The real data for use in our study is gathered mainly from our own data center on campus. We now elaborate further on our methodology.

A. Data Requirements for the DSS

Data on Thermal Profiling: An essential step in properly monitoring data center cooling costs is to gather data on thermal profiling or in other words develop a thermal profile of the data center. A thermal profile is a map of the temperature variances across the data center. To optimize cooling costs, it is important to prevent hot or cold spotting. The *terms hot and cold spotting* refer to what occurs when the temperature becomes extremely high or low respectively, outside the manufacturers' recommendation thermal conditions. Hot spotting in particular can lead to a server or rack of servers going down. In order to prevent such scenarios from occurring, a thermal profile of the data center is necessary, and wireless thermometers need to be added in potential hot spot areas. By adding wireless thermometers, data center operators can be warned when temperatures become too extreme. This will be incorporated in the DSS by using real data on thermal profiles obtained from meters in the data center.

Data on Energy Usage: One of the most important parameters to document is the amount of energy in kilowatt/hours the data center is using. In our study we measured the amount of kilowatt/hours used by manually recording the usage patterns from the Power Distribution Units (PDU). In the future starting in the Fall 2013 we will have the ability to connect remotely to record the amount of kilowatt/hours used in real time. This ability will give our research large volumes of data for mining in order find more interesting patterns.

Data on Virtualization: The concept of virtualization offers information technology administrators the ability to consolidate and optimize servers to reduce power and cooling costs. Three of the main advantages to virtualization are dealing with underutilized servers, addressing data centers running out of space, and mounting system administration costs [13]. Many servers are still underutilized running either one or a few applications. Servers should be checked to determine their utilization rates, and with virtualization software, servers can increase their efficiency rates. In addition, administrators are physically running out of space to both plug servers in, and the physical space to deploy servers. By retiring old servers, or consolidating applications using virtualization, administrators can reduce their computing power to fewer servers with a higher density rate. A possible consequence of this reduction strategy with a higher density rate is increased heat, which an administrator should check with both facilities and the hardware provider. The growth in data centers has resulted in more demands to maintain hardware and software by system administrators. At present a software provider for virtualization such as VMware is used to collect data on the virtualization rate.

Data on Carbon Footprint: Another parameter to be recorded is the carbon footprint of the data center. The carbon footprint of an organization is the estimated total of the output of carbon dioxide into the atmosphere from primarily burning fossil fuels to supply the power for in this case the data center. The formula to calculate the carbon footprint is as follows:

Carbon footprint = Electricity usage/year * U.S. CO_2 national emissions average *1 metric ton/2,204.6 lbs.

To calculate the carbon footprint of our data center we substitute in our values as follows:

Carbon footprint = 788,400 kWh/year * 1.34 lbs./kWh * 1 metric ton/2,204.6lbs = 479.21 metric tons/year

The electrical usage value of 788,400 kWh/year is a combined value of the energy use of our servers and air-conditioning system for 2010. The value 1.34 lbs./kWh is the national average of US CO_2 emissions [14]. A metric ton conversion ratio is used because CO_2 emissions are commonly expressed in the international community in metric tons.

Data on Power Usage Effectiveness (PUE)

A final parameter that is important to measure is the Power Usage Effectiveness (PUE) that measures the efficiency of power used in a data center, and is defined in the following equation*:*

PUE=Total Facility Power/IT Equipment Power

By theory a PUE of 1.0 would be perfectly efficient and there would be no power loss since the Total Facility Power would equal the draw by the IT Equipment Power. In practice, a PUE of above 2.0 is common in industry due to power loss in Total Facility Power for energy use by such components as lighting and cooling. There has been a trend by some large data centers towards a PUE of 1.0 by more efficient design factors such as applying airside economizing or also know as free cooling. Airside economizing uses outside air when factors are appropriate to cool the data center. In particular there has been a trend to locate data centers in more northern latitudes to take advantage of cooler climates, and particularly stable governments with cheap electricity such as Sweden that also has hydroelectric power that does not contribute to the carbon footprint of the data center.

B. Case Based Reasoning for the DSS

A technique that can be employed in the proposed DSS is case-based reasoning (CBR) which is the process of solving new problems based on analogous solutions of similar past problems [1]. CBR can be a powerful tool when developed in a database format with the ability to retrieve similar cases for reuse and revision. The most popular CBR cycle is the R4 paradigm, which involves a 4-step cycle of retrieve, reuse, revise and retain, i.e. retrieve the most similar existing case, reuse it as closely as possible, revise it using adaptation to suit the current case and retain the revised case for further use. This R4 cycle can be used in order to collect data on similar past cases in various green energy initiatives and draw analogies with them to provide solutions to current cases in green computing of data centers. Metrics to measure data center efficiency are being developed that can be used for comparison against other data centers for benchmarking [5, 10]. We would deploy these as needed and develop new ones specific to our problem. For data center management, the use of CBR can be supportive when updating hardware in the sense that previous cases can be obtained and the processes can become more automated, with adaptation as needed.

For instance, consider the scenario illustrated in Figure 1 with respect to Google's data center suggesting that moving the email for a large organization can lower the PUE and the carbon footprint. Drawing from the R4 cycle can be useful for administrators to retrieve previous cases where the PUE and carbon footprint have been decreased at other organizations.

The documentation of previous cases on PUE and the carbon footprint enables the administrators at Google to reuse past performance metrics to understand the changes and costs of the operation of their own data center. By drawing on previous cases and revising for the current scenario perhaps the greatest decrease in the carbon footprint to shifting to Gmail is to decrease of 103kg/year to 1.23 kg/year [22]. After testing and repairing the Google case, the knowledge is retained as a learned case and stored in the previous cases database. In this instance the PUE was reduced by 46%. The adaptation is done based on various parameters in the current situation and the learned case is stored for future use along with the relevant parameters, for use by similar other institutions. This example illustrates how CBR is useful in developing the proposed DSS in green computing, and to promote energy efficiency by saving 167kWh/year and reduce carbon emissions by up to 100 kg/year.

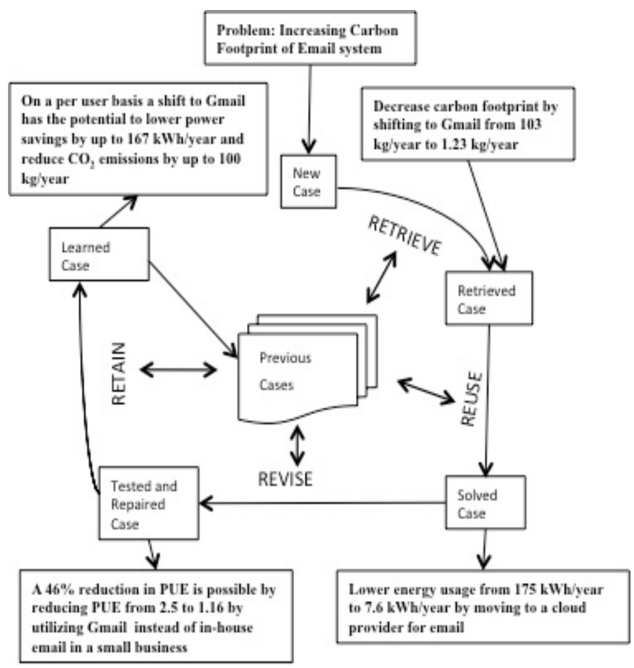

Figure 1: CBR Example for Lowering Carbon Footprint

Now consider Figure 2, where the main problem depicted is increasing operational costs. From previous cases at a Google data center it has been documented from a retrieved case that the PUE can be lowered from 2.4 to 1.5 by following three strategies from solved cases such as:

1. Optimizing air flow through better duct work and performing a thermal profile
2. Decrease server sprawl by increasing the server utilization rate
3. Raise the temperature in the data center from 72 degrees Fahrenheit to 81 degrees Fahrenheit

The result from following such strategies is that there is a 62% reduction in PUE, and from the learned case in this example a $25,000 investment led to a reduction of 670MWh/year that translates to a $67,000 savings per year [23]. This specific case makes it clear that sustainable solutions are possible for future construction and retrofitting data centers.

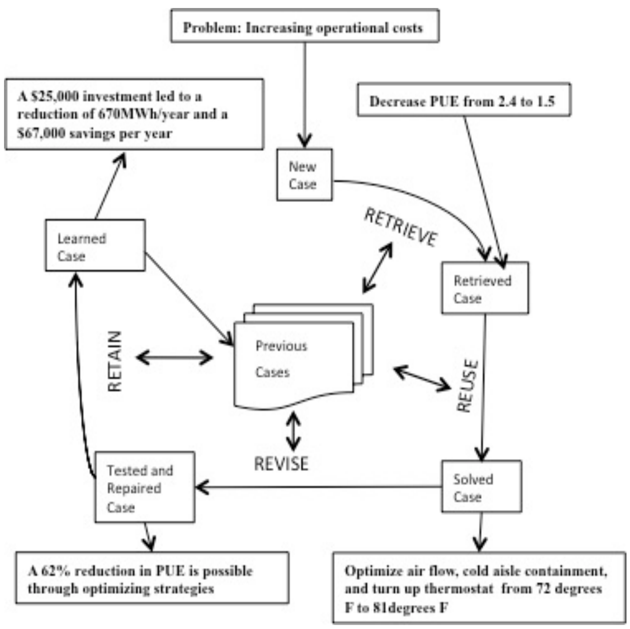

Figure 2: CBR Example for Lowering Operational Costs

Reviewing the CBR examples the common theme is lower resource use. This theme of using fewer resources is important in two aspects:

1. By using less energy the amount of carbon dioxide is reduced per year if fossil fuels are used to provide the energy. This is important for current legislation in the European Union and pending legislation in other countries to combat environmental hazards.
2. By using less materials through such factors of reducing server sprawl, virtualization and greater server utilization rates reduces the amount of processed raw materials.

These two aspects are critically important for lowering the PUE and reducing the amount of materials in data centers. Particularly the reduction in energy use is vital since the data center industry is expanding each year with the greater shift towards electronic communications and entertainment. The importance of making the data center industry more energy efficient helps mitigate the environmental hazard of climate change, and while the industry is expanding, and therefore contributing a greater percentage of carbon dioxide per year; a call for applying sustainability science is necessary for developing a DSS. Likewise, several other CBR examples have been developed which serve as useful inputs for developing a DSS for green data centers.

C. *Decision Trees for the DSS*

While the CBR paradigm entails case-specific reasoning, we would also need generic decision-making based on rules and paths that depict a logical flow, which brings us to the deployment of another data mining paradigm, namely decision trees. Decision trees provide systematic analysis based on their structure of the root representing a starting point, paths representing various decision choices and leaves representing the actual decisions or potential outcomes. This is useful as illustrated in the following examples.

A significant aspect of data centers is the utilization of servers that forms an important part of our work and will be the focus here. Data centers are fueled by the incessant demand for information and continuous uninterrupted service by users. Below is an example of a decision tree for increasing utilization rates:

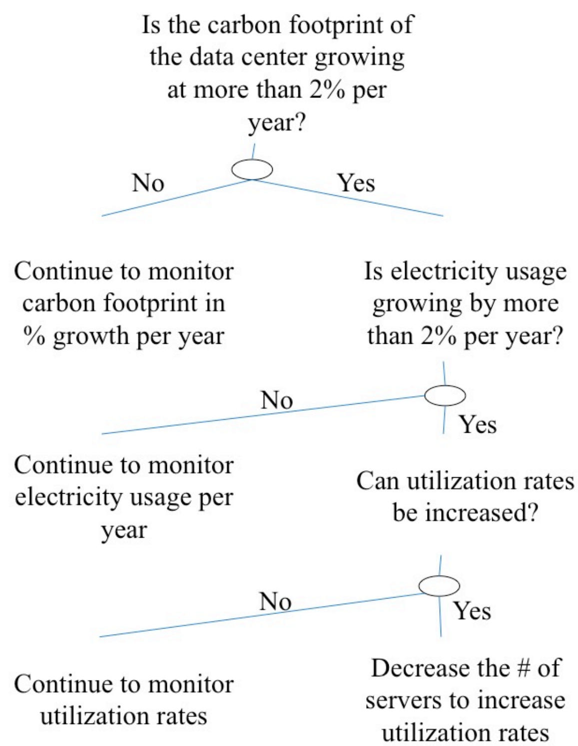

Figure 3: Decision Tree based on Server Utilization

This tree illustrates the monitoring of *carbon footprint* and *electricity usage*, both important parameters to be considered in the greening of data centers. If both these parameters show a growth of more than 2% per year, then we would try to increase the utilization rates on the servers and consequently decrease the number of servers. The 2% per year growth rate was selected as an arbitrary goal or the rate could be selected by an organization depending on the growth of the particular organization. The rate of growth would be typically higher in an organization such as a successful start-up. We selected the 2% growth rate since our study was of a mid-sized university where the data center administer was seeking to balance server sprawl, virtualization, phantom servers, energy efficiency and the carbon footprint. However, if it is not possible to increase the utilization rates then we would simply continue to monitor them. In that case, this tree per se would not yield the optimal solution but we would consider other alternatives (not shown herewith).

This policy of optimizing servers at greater utilization rates would reduce server sprawl, with the following advantages:

- Lower electricity usage that translates to cost savings and a reduced carbon footprint
- Decrease management costs to maintain and service additional servers
- Free up floor space on inefficient or phantom servers
- Provide greater efficiency in the use of resources

The decision tree in Figure 4 represents an examination of the *CPU busy rate* and *idle time*. The CPU busy rate is defined as the percentage of time during which the CPU is actually processing operations. The idle time is calculated as the percentage of time during which nothing is being processed (clearly depicting underutilization). We examine the CPU busy rate by dividing it into four equal ranges of 25% each. Following each branch of the tree gives a leaf with a recommended action or statement. For example, if the CPU busy rate is less than 25%, then a significant downsizing of the data center is recommended. The paths in this tree are self-explanatory. It is to be noted that a phantom server shown in the last leaf refers to a server whose utilization is so low that it has reached the state of a phantom or a ghost. A phantom server is a server or rack of servers that has been left on where the units are not performing any functional work. This scenario of phantom servers typically occurs over time where new system administrators take over a data center and are not sure of what every server is performing, so the servers are simply left on to use up resources.

From our study, we found that our own data center falls between the 25-50% range, and therefore depending on idle time the data center is too large. The recommendation is to increase virtualization if the idle time is greater than 75%. Virtualization is the process of placing more applications on fewer servers to increase CPU busy rate. The 75% CPU busy rate is significant since the rate suggests a trade off scenario where performance or speed is balanced between higher energy efficiency. If the idle time were less than 75%, it would be advisable to shift future demand to a cloud provider, while keeping current operations as they are. More details on this shift to cloud technology can be found in [12].

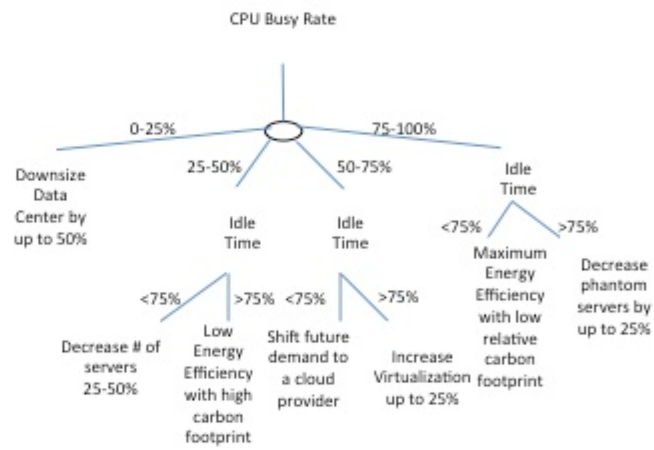

Figure 4: A Decision Tree based on CPU Busy Rate

Based on several such decision trees and CBR examples, we provide the design of the DSS. The analysis conducted through such data mining paradigms serves as the basis for implementing the DSS taking into account prospective user queries pertaining to such analysis. Users can then interact with the DSS by entering specific queries and getting responses such that they help support decision-making in data center administration with an eco-friendly perspective.

IV. DSS Questions And Answers

The following is a list of potential questions that users could pose to the DSS. We include pertinent answers to these based on our literature search, study of our own data center and analysis using data mining approaches such as CBR. Please note that a DSS is only designed to support or assist the users' decision-making, not to actually make decisions on behalf of the users. Hence these answers are estimates entailing heuristic analysis.

Q1. How can the PUE be decreased at a data center on the scale of Google?

A1. The three best ways to lower the PUE from 2.4 to 1.5 are to:

- optimize air flow
- place plastic cold aisle containments
- and increase the thermostat from 72 degrees Fahrenheit to 81 degrees Fahrenheit.

Q2. How much money needs to be invested to lower the PUE?

A2. We estimate an initial $25,000 was invested that resulted in the savings of:

- $67,000 per year or a reduction in electricity by 670MWh/year

Q3. How much is the estimated reduction in the carbon footprint when the temperature is increased by approximately 10 degrees Fahrenheit?

A3. 670,000,000MWh/year*1.34 lbs./kWh *1metric ton/2,204.6=407,239 metric tons/year

- This is an assumption that the data centers at Google are using fossil fuels; however, in reality Google is using carbon offsets to be carbon neutral.

Q4. How much can you reduce the carbon footprint of a small business's email system by switching to a cloud provider for email service?

A4. It has been estimated that a small business with two servers for email uses approximately:

- 400 Watts per year that translates to 8 Watts per year per user
- The carbon footprint can therefore be reduced from103kg per year to 1.23 kg per year using a cloud provider

Q5. What is the reduction in annual power per user by switching from a small business email system to a cloud provider?

A5. The energy usage can be reduced from an estimated:

- 175 kWh per year to 7.6 kWh per year

Q6. What is the reduction in PUE in numbers and percent from adopting this strategy of shifting to a cloud provider for email service?

A6. The PUE can be reduced from 2.5 to 1.16 that results in a 46% reduction.

Q7. What is the overall savings in electricity and carbon emissions by shifting a small office?

A7. By selecting a cloud provider such as Gmail an office can save up to 167 kWh per year and lower carbon emissions by up to 100 kg/year.

Q8. What is the reduction in PUE by shifting customer relationship software to a cloud provider such as Salesforce.com?

A8. A shift in strategy to deploying to the cloud can lower the PUE from 1.97 which is an industry average to 1.53 which is the average PUE for Salesforce.com

Q9. What is the average reduction of carbon emissions by shifting to customer relationship software that is hosted in the cloud?

A9. The range in reduction is from 80-95% per user by shifting to a cloud provider for software services.

Q10. How many tons per year of CO_2 emissions did customers who switched to a cloud provider such as Salesforce.com eliminate from entering the atmosphere?

A10. By switching to a cloud provider an estimated 170,900 tons/year of CO_2 emissions were eliminated from entering the atmosphere by using cloud software.

V. RECOMMENDATIONS FOR USERS

The key to gaining the greatest energy efficiency to lower environmental hazards and to promote sustainability in data centers is a combination of a few strategies such as the following:

- Promote higher server utilization to prevent server sprawl by placing more applications on less servers
- Modern servers do not need to be kept as cool as previous generation main frame computers, so increasing the temperature in the data center will not damage equipment and result in significant financial savings (as demonstrated in Figure 2.)
- Move applications better suited for the cloud such as email to a cloud provider such as Google that has greater economies of scale that will result in increased services with lower cost and a lower carbon footprint.
- Perform thermal profiles to audit cooling waste in order to promote better air flow
- Continue to virtualize the data center which will result in less servers operating and the resulting decrease in cooling from less servers

Following the above strategies will be a first step towards promoting a more sustainable world. The result will be significant cost reductions with lower energy use that translates to a lower carbon footprint if fossil fuels are used to power the data center. Our research team believes that by promoting our work plus the work of others will result in reducing environmental hazards.

VI. RELATED WORK

Green technology and environment-friendly products have been studied from various perspectives, especially due to their contributions towards a sustainable planet. ASHRAE [3] provides recommendations for specific parameters to be adhered to in order to stay within safe ecofriendly limits. The Green Grid [5] describes performance metrics from the specific angle of data centers. These are helpful sources that provide guidelines to several researchers and practitioners.

Daim et al. also propose data center metrics along with a model for energy efficiency. The work of Matthew et al. [9] explains some benchmarks and recommended actions for data centers. Stanley et al. [15] quantify the problem of data center greenness and energy consumption with a view towards profitability. Best practices in data center usage are outlined in the works of Tschudi et al. [18].

All these sources are useful in our literature survey and have been valuable references in our research. In addition, works such as those of Aamodt et al. [1] in CBR, Ambrust et al. [2] in cloud computing, Clark et al. [6] and Palmer et al. [11] in sustainability have also provided certain inputs for our study. White papers on green computing and related areas appear in [10], [15], [16] [20], [23] and [24], thereby highlighting the importance of this research in the 21^{st} century for protecting the environment.

To the best of our knowledge, ours is among the first works to perform data mining guided by domain knowledge in environmental management in order to build a DSS for green data centers. That is a novel contribution on our part. The DSS would be useful for data center managers and also to professionals in data mining, green computing and environmental management.

VII. Conclusions And Future Work

In this paper, we have presented a case study on the greening of data centers. We have discussed the data requirements, depicted the paradigms of CBR and decision trees with real examples for DSS development, outlined potential DSS questions and answers, and made recommendations to users accordingly. In the past the data for this study has been collected manually. This approach to data collection has proved useful to gain an idea of how the system works. However, a steady and continuous data stream is necessary for data mining to find useful patterns for energy efficiency. Therefore, the next stage of this study will include metering for energy usage of the data center to predict such factors as the PUE and to find trends in the data for energy efficiency. In the summer of 2013 these meters and software are being installed for an enriched study. The research team plans to collect and analyze these data streams over the 2013-14 academic semesters.

In the future the research team plans to enhance the DSS design and implement the full-fledged system. This DSS would be helpful in the greening of data centers by minimizing server sprawl, lowering carbon footprint and so forth. It would thus be an important step towards data center sustainability, having a positive effect on the environment. Hence, the broader impacts of this work would involve reduction in environmental pollution and streamlining climate change. We therefore address important environmental hazards in our work with a view to minimizing them. The analysis, design and implementation pertaining to the DSS are useful contributions in this process.

VIII. Acknowledgments

The authors would like to thank the PSEG Company in NJ for a research grant that provides the funds for the major part of this activity as part of their green technology initiative. This work was also partly supported in its inception by funds from the Dean of the College of Science and Mathematics at Montclair State University in an effort to green the campus.

IX. References

[1] Aamodt, A., and Plaza E. (1994) "*Case-Based Reasoning",* Artificial Intelligence Communications, IOS Press, Vol. 7: 1, pp. 39-59.

[2] Armbrust, M., Fox, A., Griffith, R., Joseph, A. D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stocia, I, & Zaharia, M. (2009) *"Above the Clouds: A Berkeley View of Cloud Computing."* Technical Report, Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA.

[3] ASHRAE. (2008) *"2008 ASHRAE Environmental Guidelines for Datacom Equipment- Expanding the Recommended Environmental Envelope."* ASHRAE White Paper.

[4] Bean, J. and Dunlap, K. (2008) "*Energy-Efficient Data Centers: A Close-Coupled Row Solution."* ASHRAE Journal 50 (10), pp. 34–42.

[5] Belday, C., Rawson, A., Pflueger, J., and Cadet, T. (2008) "*Green Grid Data Center Power Efficiency Metrics: PUE and DCIE."* Green Grid, White paper # 6.

[6] Clark, W. C. and N. M. Dickson (2003). "Sustainability science: The emerging research program.", Proceedings of the National Academy of Sciences of the United States of America, vol. 100, no. 4, pp. 8059-8061.

[7] Daim, T., Justice, J., Krampits, M., Letts, M., Subramanian, G. and Thirumalai, M. (2009) "*Data Center Metrics: An Energy Efficiency Model for Information Technology Managers*." Management of Environmental Quality: An International Journal, Vol. 20, pp. 712 -731.

[8] Koomey, J. (2007) *"Estimating Total Power Consumption by Servers in the U.S. and the World."* Technical Report, Stanford University, CA.

[9] Matthew, P., Ganguly, S., Greenberg, S., and Sartor. D. (2009) "*Self-Benchmarking Guide for Data Centers: Metrics, Benchmarks, Actions."* Technical Report, Berkeley Lab, CA.

[10] NRC (National Research Council) (2002). Our common journey: A transition toward sustainability. Washington, DC: National Academy Press. [White Paper]

[11] Palmer, M., E.S. Bernhardt, E.A. Chornesky, S.L. Collins, A.P. Dobson, C.S. Duke, B.D. Gold, R.B. Jacobson, S.R. Kingsland, R.H. Kranz, M.J. Mappin, M.L. Martinez, F. Micheli, J.L. Morse, M.L. Pace, M. Pascual, S.D. Palumbi, O.J. Reichman, A. Simons, A.R. Townsend, and M.G. Turner. (2005). Ecological science and sustainability for the 21st century. Frontiers in Ecology and the Environment 3: 4–11.

[12] Pawlish M., Varde A., Robila S. (2012) *"Cloud Computing for Environment-Friendly Data Centers"* ACM CIKM CloudDB workshop, pp. 43-48.

[13] Schmidt, R., Beaty, D., and Dietrich, J. (2007) "*Increasing Energy Efficiency in Data Centers."* ASHRAE Journal, December, pp. 19-26.

[14] Schulz, G., 2009. "*The Green and Virtual Data Center",* Auebach Publications, Boca Raton, FL.

[15] Stanley, J., Brill, K., and Koomey J. (2007) "*Four Metrics Define Data Center 'Greenness': Enabling Users to Quantify Energy Consumption Initiatives for Environmental Sustainability and "Bottom Line" Profitability."* Uptime Institute Inc., [White Paper]

[16] Sun Microsystems. (2009) "*Energy Savings in the Datacenter."* Sun Microsystems Inc. [White Paper]

[17] Talebi and Way (2009) "*Methods, metrics and motivation for a green computer science program.",* ACM SIGCSE, pp. 362-366.

[18] Tschudi, W., Mills, E., Greenberg, S., and Rumsey, P. (2006) "*Measuring and Managing Data-Center Energy Use: Finding and Resulting Best Practices—From a Study of Energy Use in 22 Data Centers."* HPAC Engineering Report 45, Marcy.

[19] United States Environmental Protection Agency (2007) *EPA Report to Congress on Server and Data Center Energy Efficiency,* Technical Report, Environmental Protection Agency, Washington, D.C.
[20] Varde A., Robila S. and Weinstein. M., (2011) *"Green Data Centers for Sustainability"* NIST-TIP: National Institute for Standards and Technology - Technology Innovations Program. [White Paper]
[21] Weinstein, M.P., R.C. Baird, D.O. Conover, M. Gross, J. Keulartz, D. K. Loomis, Z. Naveh, S.B. Peterson, D.J. Reed, E. Roe, R.L. Swanson, J.A.A. Swart, J.M. Teal, R.E. Turner, and H.J. van der Windt. (2007). "*Managing coastal resources in the 21st century"*. Frontiers in Ecology and the Environment 5: pp. 43–48.
[22] Witten, I. and Frank E. (2000) "*Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations"*, Morgan Kaufmann Publishers.
[23] Google (2011) "*Google's Green Computing: Efficiency at Scale*." Google [White Paper]
[24] Google (2011) *" Google's Green Data Centers: Network POP Case Study"* Google [White Paper]