A Decision Support System for Green Data Centers

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

ABSTRACT

In this paper, we propose a decision support system for green computing in data centers on campuses. Green computing aims at the development of technologies for a greener and more sustainable planet. In this work we focus on the greening of data centers that house servers on campuses. It is known that servers consume a huge amount of power and incur other costs in cooling and related operations. It is challenging to meet the demands of efficiency and accuracy in data centers while yet promoting a greener environment along with cost cutting. Our work proposes a decision support system based on decision trees and case-based reasoning that mines existing data in order to assist decision-making for better management in data centers. We consider various aspects of the data from an environmental management perspective such as carbon footprint, thermal profiles and virtualization. Issues such as efficiency and accuracy are taken into account here, In this proposal paper, we outline the need for our work, the challenges involved, the proposed approach for building the system, preliminary evaluation and future plans. This work would be of interest to the data and knowledge management community as well as environmental scientists and researchers in related areas.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications – *data mining;* J.2 [**Computer Applications**] Physical Sciences and Engineering – earth and atmospheric sciences.

General Terms

Management, Measurement, Performance, Design, Experimentation, Human Factors.

Keywords

Case-Based Reasoning, Data Centers, Decision Trees, Decision Support Systems, Green Information Technology.

PIKM'10, October 30, 2010, Toronto, Ontario, Canada.

1. INTRODUCTION

Humankind is not living sustainably, and action must be taken, and taken soon, to reverse this negative trend and reconcile human use of the planet's resources with the ability of the earth to supply them. A successful sustainability transition, moreover, will demand critical advances in basic knowledge, in humankind's social and technological capacity to utilize it, and in the political will to turn that knowledge and know how into action [11]. The process goes beyond individual stakeholders and themes – populations, economy, water, food, energy, and climate – to identification of common threads and drivers of systemic change [11]. 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]. In 2006, the US Congress passed a law (P.L. 109-431) that required the USEPA to submit a report on data center energy consumption. Similarly, the European Union issued a voluntary "code of conduct" in 2007. 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, e.g., energy usage by the nation's servers and data centers is significant, and estimated to be about 61 billion kilowatt-hours (kWh) in 2006 at a total electricity cost of about $4.5 billion. This estimated level of electricity consumption is more than the electricity consumed by the nation's color televisions and similar to the amount of electricity consumed by approximately 5.8 million average U.S. households.

The energy use of the nation's servers and data centers in 2006 was estimated to be more than double the electricity that was consumed for this purpose in 2000. Under current efficiency trends, national energy consumption by servers and data centers could nearly double again in another five years (i.e., by 2011) to more than 100 billion kWh, representing a $7.4 billion annual electricity cost. The peak load on the power grid from these servers and data centers is currently estimated to be approximately 7 Giga-watts (GW), equivalent to the output of about 15 base-load power plants. If current trends continue, this demand would rise to 12 GW by 2011 [19]. Thus the development of green energy initiatives for sustainability in data centers is clearly of global significance. Green energy was one of the topics mentioned in the State of the Union address of President Obama and has been identified as an item of national priority.

The proposed research is in the area of greening 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.

2. PROBLEM DEFINITION

2.1 Goals

The main goal of this research is to conduct a detailed analysis of data centers from a green 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

2.2 Novelty

This study would be one of the first for decision support in green data centers. The benefits of the proposed study would have positive implications for data center system administrators by offering detailed case studies of reductions in electrical and carbon emissions. A majority of case studies in the literature conducted so far, have been through vendor companies that are not peer-reviewed. By providing a third party audit of a data centers, we will aim to increase the dissemination of knowledge in the efficiency of data centers, for the benefit of various users all over the world.

2.3 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 electric meters on the air-conditioning system in the near future, which presents a challenge in data collection.

An additional challenge to implementing a green computing program in academia is the need to bring different disciplines together. There is a need to draw together the disciplines of computer science, environmental management and related departments such as the operations of information technology and facilities. It is challenging to bring together experts in their respective fields with various levels of knowledge on data center issues, which is not an easy task, keeping in mind the pressing needs of professionals to cater to other priorities

3. PROPOSED APPROACH

We propose to build the required decision support system for green data centers using a data mining approach comprising decision trees and case-based reasoning, 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 at Montclair State University (MSU). We now elaborate further on our methodology.

3.1 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, hot or cold spotting is important to prevent. Hot and cold spotting occur when the temperature becomes too extreme or 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 this scenario 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.

Data on Energy Usage: One of the most important parameters to document is the amount of energy in kilowatts the data center is using. In our study we measured the amount of kilowatts used by manually recording the usage patterns from the Power Distribution Units (PDU). For future data centers, we would recommend that the ability to track energy usage be connected remotely to record the amount of kilowatts used in real time. Documentation of energy usage is critical to track changes in servers, develop efficiency metrics, and virtualization.

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. A virtualization strategy that reduces the number of servers will also cut costs on the number of system administrators needed, or the time to maintain the servers.

Data on Carbon Footprint: Another parameter to be recorded is the carbon footprint of the data center. The carbon footprint measures the output of carbon dioxide into the atmosphere. It is important to monitor and control this because it is desirable to keep the carbon footprint low for heading towards an environment-friendly data center.

Data on Relative Humidity: The data center room temperature is typically set to a steady 68 degrees Fahrenheit with slight variations which also needs to be measured. Inside the data center the relative humidity levels need to be recorded. Too high of a relative humidity level poses a problem of moisture on the equipment, and too low of a relative humidity level can cause static electricity problems. Thus, the outside temperature for both the high and the low need to be recorded, as well as the humidity level with respect to the given data centers.

Data on Acoustic Levels: Data centers are loud due to the cooling and fans needed to keep the servers at an optimal temperature. Increasing the temperature to save on cooling costs can result in increased noise due to the server fans operating more frequently. The resulting noise presents a hazard to workers due to the increase in the decibel level, and also presents a challenge when communicating on a phone due to the background noise. While we have not measured the decibel level of the data center at MSU, we feel that it will be increasingly important to measure this parameter. We recommend that data centers install wireless remote decibel meters to gain knowledge on operations in order to measure any changes, and for the safety of personnel. Remote sensors that measure the decibel level would provide additional data on the performance of a data center that would be important to document.

3.2 Techniques for the DSS

Decision Trees: As widely known to data mining professionals, decision trees are a stem and leaf figure to represent possible outcomes where the root represents the starting point, the branches represent various paths or alternatives considered and the leaves represent the final decisions. At the core of the decision tree is the assignment of probabilities of possible outcomes to assist in complex decision-making [21]. As data center managers would corroborate, the many decisions needed to build or update a data center can be overwhelming. By using decision tree analysis, the large number of different factors can be taken into account and thereby reduce the uncertainty.

Figure 1 shows a small snapshot of a decision tree developed to calculate the net present value (NPV) of energy costs when deciding between a new purchase of either Energy Star servers or standard servers. At the heart of the performance versus cost issue when purchasing new servers are the new tier of servers that are Energy Star rated by the U.S. Environmental Protection Agency [19]. An Energy Star appliance, or in this case server, makes the decision processes a little easier; however, since most data centers have a mix of hardware suppliers implemented in their data center, it will still be a major decision to settle on which vendor to select. Using arbitrary values for now, a decision tree is developed in Figure 1 by starting with two products and showing the initial cost of both, and then based on the net present value of energy cost over time; the decision tree would demonstrate the total operating cost over the product life cycle. For this example, it is estimated that an Energy Star server costs 20% more than a standard server, however different scenarios are presented on energy costs over time, and the key aspect of this example is to examine the total cost of ownership over time. The result of

Figure 1 is that the total cost of ownership over the three year period varies from an optimistic scenario of $28,000 to a historic scenario of $40,000. The conclusion from this scenario would be the recommendation to purchase Energy Star servers from a purely financial perspective.

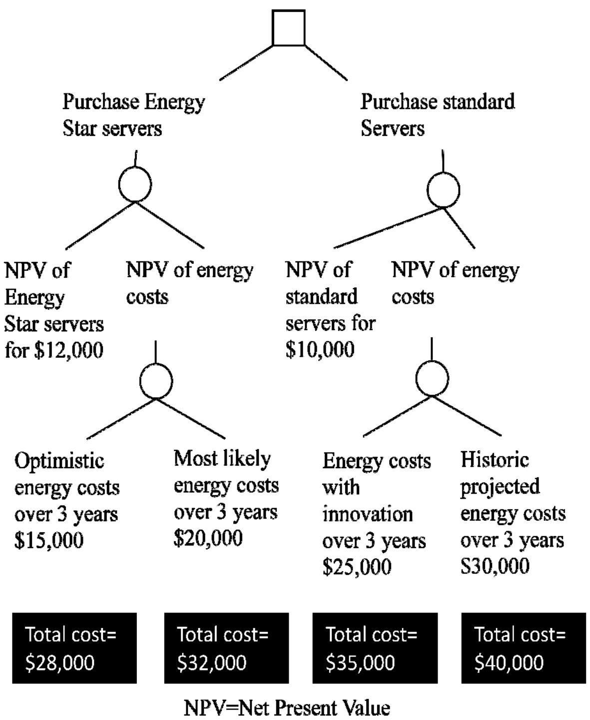

Figure 1: A Decision Tree Example

Decision trees can also assist the data center management team in the area of virtualization. Currently it is estimated that most servers are running at about 25%-35% utilization rates and should be approaching a utilization rate of closer to 85% [4, 16]. In addition, there are often servers that are phantom servers that are left on from previous applications that are not used any more. A decision tree can take stock of all the servers and racks, and assist in the virtualization to better manage the data center. By both eliminating phantom servers and consolidating servers, there are numerous benefits such as reduced electric bills, reduction of CO_2 emissions, freeing up of valuable floor space, and lower administration costs [13].

Case Based Reasoning: Another 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.

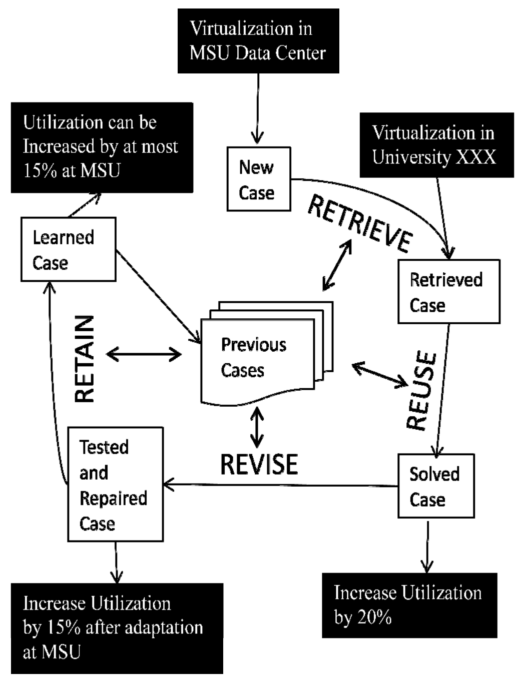

Figure 2: The CBR R4 Cycle with an Example Case

For instance, consider the scenario illustrated in Figure 2 with respect to Montclair State University (MSU), considering a case of virtualization virtualization, using arbitrary numbers for now. The popular CBR R4 cycle in the literature, e.g., [1] is useful here. Servers have a life cycle of approximately three to four years, and applying the CBR cycle when updating for virtualization is important to properly and efficiently configure the new data center. Drawing from the R4 cycle can be useful for administrators to retrieve previous cases where utilization rates have increased in other universities. In this case, let us consider University XXX. The documentation of previous cases on virtualization there enables the administrators at MSU to reuse past performance metrics to understand the changes and costs of the operation of their own data center. When replacing or updating servers, virtualization software enables the administrator to reduce the number of servers by increasing the utilization rate of the new servers. It is important at server replacement time to examine the solved cases, and for this example there was an increase in utilization at University XXX by 20%. By drawing on previous cases and revising for the current scenario perhaps the greatest increase in utilization after adaptation at MSU is 15%. After testing and repairing the MSU case, the knowledge is retained as a learned case and stored in the previous cases database. 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.

Combination of Approaches: CBR would be used in our DSS conjunction with decision trees, as justified by the following logic. There are some scenarios that can be better addressed by comparison with existing ones thus opting in favor of a CBR approach, while some that can be more easily solved by applying standard rules and regulations or approximations thereof, represented by tree paths, thus suggesting a decision-tree approach. We would consider the use of both in certain scenarios and compare the results, leaving it to the domain experts to determine the better outcome and tuning that further if needed to an optimal solution, which would also be stored in a database for future use. In some scenarios, we could actually accept the results from both the approaches, rather than choosing one of the two, because each approach would yield some additional knowledge enabling better decision-making. Our deployment of decision trees and CBR would thus be guided by domain knowledge and the resulting solutions would be reusable. Further research on individual sub-problems that may arise is ongoing and more challenges will be addressed as they are encountered in development.

Potential Use of Cloud Computing: Experts in the field of data management are claiming that the time for using cloud computing has arrived. We find considerable research on cloud computing in data management conferences and journals. Cloud computing is a nebulous term; however, as Armbrust [2] state cloud computing serves as a utility and has the possibility to transform the information technology industry. Companies can now build a company without the capital expense of purchasing a data center, and there is also not a need to plan for future expansion of the data center. This data-center-less scenario has the advantages of purchasing only what an organization needs in a data center, and system administrators do not have to worry about under-provisioning or over-provisioning for data center needs. Under-provisioning would be when the data center is not built up to handle the traffic generated, and over-provisioning is when the data center is being underutilized. Presently, many data centers are being underutilized due to data centers being built for maximum capacity which never materialize or only have small spikes throughout the year.

Cloud computing can turn a capital expense into an operating expense which will lower the initial costs of starting a business or expanding a data center. Data centers can now grow using cloud computing as a utility. Large data centers or data center farms located where there is cheap electric power can fuel this growth. Therefore, cloud computing offer the flexibility and scalability to data center operators when expansion is needed without the need for estimating future demand for services.

4. SYSTEM IMPLEMENTATION

We give an overview of the general system architecture of our proposed decision support system for green data centers. This is followed by preliminary data analysis and further experimental plans.

4.1 System Architecture

A broad outline of the system architecture of our DSS is presented here. We collect input data pertaining to parameters such as temperature, humidity, energy usage and acoustic levels. We aim to conduct detailed analysis using data mining techniques to discover knowledge for achieving more efficient operations of data centers.

As stated earlier, we propose to employ the data mining approaches of decision trees and case-based reasoning to analyze various parameter adjustment scenarios, and their possible outcomes with respect to greener computing. This general system architecture is summarized in Figure 3. It is a very high level architecture diagram summarizing the big picture. Expanding on the details of its component constitutes part of our ongoing work in system implementation.

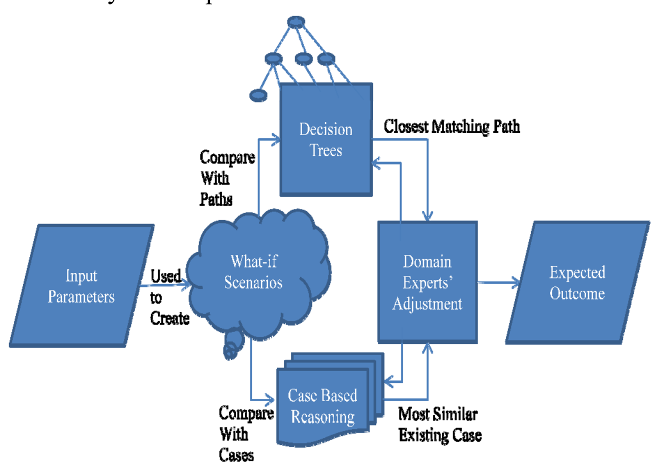

Figure 3: General High-Level Architecture of DSS

In this figure, the input parameters refer to various aspects such as temperature, humidity, energy usage and acoustic levels as stated earlier. The process of data collection is manual as of now, and we aim to partially automate it with the buying of new equipment. The what-if scenarios indicated here represent *what* the data center professionals would be interested in analyzing to envisage an outcome *if* a particular step was taken. As an example, a what-if scenario could be *"what would be the consequences if we raised the temperature in the data center room by x degrees"*. The use of case-based reasoning and decision trees for analysis has already been explained earlier and the boxes in this figure represent the corresponding modules for development. The expected outcome in the figure is what we would see as the result of the what-if scenario that assists in decision-making. The domain experts' adjustment shown here represents the human interpretation of the data with specific adjustments as needed in order to achieve a more meaningful expected outcome. This would be stored in a database for future use. There is also a feedback loop in our architecture based on this expert adjustment, due to which we can further analyze the scenarios using decision trees and CBR.

The system development of the DSS would be conducted using Java. The well-known Waikato Environment for Knowledge Analysis (WEKA) [21] would be used for conducting some of the data mining tasks. A good user interface would be developed for interaction with targeted users of the DSS.

4.2 Preliminary Data Analysis

Preliminary data has been collected in the data center in University Hall at Montclair State University in New Jersey. This data center serves the entire campus and is representative of data center operations in several universities. We have been recording this data since July 2009. Students from various departments in the school such as business and humanities have been involved with the data collection as this is a campus-wide initiative supported by the offices of the respective deans.

The data collected depicts energy use of the servers, inside and outside temperature, outside humidity and relative humidity of the server room. Figure 4 shows the trend of energy use of the servers. The servers in the data center are on four power distribution units (PDUs). The chart shown in Figure 4 demonstrates that PDU 2 is being underutilized and that PDU 3 is being used the most. Besides the month of December, there appears to be an upward trend in all the PDUs which could be attributed to the addition of servers in the data center room.

Given this PDU data, the carbon footprint was calculated as follows. The national average in the USA for emissions of carbon dioxide was used, which is estimated to be 1.34lb/kWh of electricity [14].

For the carbon dioxide output = 351,575 kWh x 1.34 lbs/kWh of CO_2/2000lbs * 2 (since we had 6 months of data) =472 tons of CO_2 per year for just the energy used by the servers (See Figure 5).

For the cooling of the servers the estimated electrical usage is 58 kW per hour with three air conditioning units running seven days a week, and 365 days per year. Thus, the equation is 58 kW*3 A/C units*24 hrs/day* 365 days year = 1, 524,240 kWh a year. To calculate the carbon dioxide output using the equation from the previous paragraph, 1, 524,240 kWh x 1.34 lbs/kWh of CO_2/2000lbs=1021 tons of CO_2 per year for the cooling of the data center.

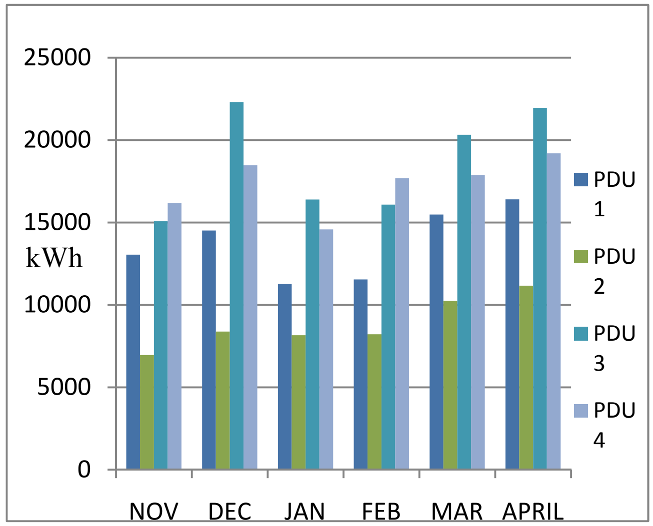

Figure 4: Energy Usage in Power Distribution Units

In order to calculate the estimated electrical cost, we used the average cost of electricity which is .14 cents/kWh of electricity (personal communications with Jim Cassidy, Head Electrician Montclair State University, 2010). Thus, the estimated electrical cost per year for the servers= 351,575 kWh * 2 (since we had 6 months of data) x .14 cents/kWh of electricity= $ 98,441. Please refer to Figure 5. Likewise, the estimated Electrical Cost per Year for cooling system= 1, 524,240 kWh x .14 cents/kWh of electricity= $ 213,394

	Total kWh Usage from 11/1/09 to 5/1/10	Estimated Total Carbon Dioxide Output per year (in Tons)	Estimated Electrical Cost per Year for the Servers
PDU 1	82269	110	$ 23,035
PDU 2	53135	72	$ 14,878
PDU 3	112144	150	$ 31,400
PDU 4	104027	140	$ 29,128
Total	351575	472	$ 98,441

Figure 5: Carbon Dioxide and Electrical Cost Estimates

4.3 Experimental Plan

Once the decision support system is developed, the focus will be on finding ways to make the data center more energy efficient. Initially the study will be focused on the campus of Montclair State University (MSU); however, the study would be open to work with other data centers in the area. In the meanwhile, as a next step, a thermal profile is scheduled to be performed in the Fall of 2010, and based on the results of the study there will be a better understanding of the cooling system of the data center. If the data center is being overcooled, administrators working with facilities can experiment with raising the temperature in the data center. Raising the temperature could be a potential area of significant savings.

Another avenue to experiment with is the use of free cooling. Free cooling is using the natural outside temperature to cool the data center after the air has been filtered. Free cooling is increasingly being used in the new construction of data centers around the world. While free cooling would be quite expensive to retrofit University Hall on our campus due to the data center being located in the center of the building with no outside wall, we would recommend that free cooling be explored for the data center that is currently being designed for College Hall. The climate of New Jersey should allow free cooling for at least six months of the year, and result in a significant return on investment in cost savings for cooling.

Detailed experiments would be performed taking into account various decision-making scenarios useful to data center professionals. Students, faculty and the administration would be involved in terms of conducting surveys at different stages of this work. Their user satisfaction would serve as a means to determine the effectiveness of this study from an accuracy and efficiency perspective. Satisfaction from a green energy perspective would also be taken into account by incorporating comparative studies before and after making the changes with respect to energy consumption and other aspects pertaining to greening the campus. We hope that this effort will yield greener and more sustainable campuses, an important step towards greening the planet as a whole.

5. RELATED WORK

Other areas of development in the field of data center operations are the improvement of data center metrics. Experts in the field have been trying to develop and improve metrics similar to the Corporate Average Fuel Economy (CAFE) standards in the automotive sector for similar efficiency standards for data centers. The purpose of introducing data center metrics, such as Power Usage Effectiveness (PUE) and Data Center Productivity (DCP) is to have comparable metrics across data centers [7, 9]. Some of the organizations that are working to develop the metrics include such groups as the Uptime Institute, Green Grid, U.S. EPA Energy Star, and Climate Savers [15]. These groups are seeking to implement the metrics across sectors to have a few equivalent standards so that there are common comparisons.

To the best of our knowledge our work constitutes a pioneering direction in the field. The strength of our project is that it constitutes work on managing a data center from a sustainability perspective, with the focus being on consumption measurement. In the market place companies such as Google, IBM and Sun have performed studies to reduce cost and green operations, but have been limited to commercially produced white papers. Related research groups could also be other universities that are implementing sustainability IT projects. At this time it is hard to gauge the extent of detail of various universities in their efforts on green IT. For example, Tufts University has been experimenting with different cooling options for their data center, but has not released an official report.

At Villanova University, Talebi and Way [17] examined green computing, however their focus was on energy savings of individual computers, and their study did not focus on the data center. The superiority of our project is that the study is conducted in house and from a cross disciplinary research team. The results will be applicable to educational facilities as well as small to medium size business units.

Server rooms are notorious for needing to be kept cool. With cooling costs of data centers representing up to 50% of the operational costs, administrators are increasingly looking for ways to reduce costs [18]. While computers still need to be kept cool, server designers are building newer product lines to have the capacity to operate effectively at higher temperatures [3]. Recently, in 2008, the American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE) widened the range of recommended temperature of data centers from 68 to 77 degrees Fahrenheit in 2004, to the current 64.4 to 80.6 degrees Fahrenheit [3]. Raising the temperature of the air-conditioners can lower operating costs; however, data center managers must check the warranty of their server contracts. Server manufacturers only warranty their products on the condition that the server rooms do not rise above a certain temperature for a specific period of time.

While the above researchers and agencies are focused on such topics as metrics, virtualization and cooling of data centers, our project aims to make even more technological contributions in that we will consider all of these topics and some more for a real university setting and moreover, from a cross-disciplinary perspective. The repercussions of our work will impact data center management resulting in a DSS to serve as a novel software development toolkit that can be deployed in similar organizations globally.

6. CONCLUSIONS

In closing, the proposed research will incorporate a decision support system (DSS) using decision trees, case-based reasoning and environmental management aspects such as thermal profiling to provide information for better decision-making in building greener campuses. To the best of our knowledge, ours would be the first DSS in green computing, which is an area of critical national need. The contributions of this work include:

- Proposing an approach based on the data mining techniques of decision trees and case-based reasoning and several domain-specific aspects in environmental management to build a DSS in green computing.
- Investigating the use of cloud computing, a hot topic for data and knowledge management professionals today, with respect to providing better alternatives in data centers.
- Working towards enhancing the operations of data centers, taking into account accuracy and efficiency as well as green energy initiatives, and balancing them well
- Addressing various challenges outlined herewith and solving further sub-problems in this research that would be encountered in the development of the individual modules of the DSS
- Conducting multidisciplinary research across the computer science fields of data management and knowledge discovery in addition to the physical science fields of environmental studies and its related areas, encouraging further such research
- Taking an important step towards sustainable technologies in computing for a greener planet, a significant broader impact.

Ongoing work includes proposing a detailed algorithmic framework for the DSS, addressing the various challenges and sub-tasks, further enhancing the proposed solution as needed, developing the required software toolkit and conducting exhaustive experimental evaluation. The results will hopefully lead towards a next generation of data centers that are more energy efficient, and provide safe working conditions for employees, while balancing with the needs of accuracy and efficiency in computing. It is expected that in this new decade, data centers will be deployed that lower the adverse impact on the environment, while still meeting the increasing demands of society. Thus, these data centers of the future would head towards green computing for a cleaner environment.

7. ACKNOWLEDGEMENTS

This research has been supported by a campus-wide initiative at Montclair State University through funds from the College of Science and Mathematics, the School of Business, the College of Humanities and Social Studies and the Office of Information Technology (OIT). We gratefully acknowledge the inputs of Dr. Stefan Robila, head of the Technical Advisory Committee in this work and Dr. Michael Weinstein, head of the Institute for Sustainability Studies at Montclair State University. We also convey our sincere thanks to Jeff Giacobbe from the OIT and students Emily Bisnauth and Yan Chowdhury for their efforts in data collection. Finally, we thank our Dean, Dr. Robert Prezant, for initiating this activity and supporting it at various stages.

8. 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."* Electrical Engineering and Computer Sciences, University of California at Berkeley.

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

[4] Bean, J. and Dunlap, K. (2008) "*Energy-Efficient Data Centers: A Close-Coupled Row Solution."* ASHRAE: October.

[5] Belday, C., Rawson, A., Pfleuger, J., and Cader, T. (2008) "*Green Grid Data Center Power Efficiency Metrics: PUE and DCIE."* Green Grid.

[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."* Stanford University.

[9] Marwah, M., Sharma, R., Shih, R., and Patel, C. (2009) "*Data Analysis, Visualization and Knowledge Discovery in Sustainable Data Centers."* ACM.

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

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

[12] 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.

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

[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.*" Association for Computing Machinery: Special Interest Group on Computer Science Education, 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: Marcy.

[19] United States Environmental Protection Agency (2007) *EPA Report to Congress on Server and Data Center Energy Efficiency,* Environmental Protection Agency, Washington, D.C.

[20] 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.

[21] Witten, I. and Frank E. (2000) Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations, Morgan Kaufmann Publishers.