Free Cooling: A Paradigm Shift in Data Centers

Michael Pawlish*and Aparna S. Varde+
*Department of Earth and Environmental Studies, Email: pawlishm1@mail.montclair.edu
+Department of Computer Science, Email: vardea@montclair.edu
Montclair State University, Montclair, NJ, USA

***Abstract*--The crossroads of sustainable development and data centers are examined due to the growing demand for electricity, and the increasing size of the carbon footprint of data centers worldwide. Free cooling involves using the natural climate to cool the data center as opposed to the more traditional method of using conventional systems such as air-conditioning. In our research, we consider retrofitting a data center as a complex decision-making problem. Case based reasoning and decision trees are two widely used techniques in the area of decision support. We propose to employ these two techniques in the framework of sustainable development to assist decision makers in the evolving design of data centers, with specific reference to free cooling.**

Keywords: Case based reasoning, Decision support systems, Decision trees, Free cooling, Green computing, Sustainability

I. Introduction

Sustainable development has been defined as the ability to provide for current generations without compromising future generations. While sustainable development has traditionally not been associated with data centers, we claim that the link is the carbon footprint of the growing number of data centers worldwide needed to fuel the mounting demand in our digital age. The demand for energy usage for data centers has been increasing sharply over the prior decade primarily for electricity needed to power and cool the data centers of the world [1]. This current demand for information has greatly benefited society. However, when viewed from a sustainable development angle, the primary concern has been the increase in greenhouse gases from power sources of non-renewable resources such as coal.

In this paper, we emphasize that as we enter a new decade, the latest paradigm shift in data center design and construction is free cooling. The shift towards free cooling stems from basic economics and data center evolution. Free cooling is using the natural temperature and the prevailing winds of the local area to cool the data center, and this differs from the traditional method to cool the data center using air-conditioning. The costs to cool data centers using conventional methods have been escalating with data center growth. Leading large scale data center operators such as Google, Intel and Yahoo have been pushing the technological boundaries of their equipment in data centers [2]. The evolution of data centers is therefore moving towards more efficient energy usage while reducing the carbon footprint of individual data centers. However, the number of data centers worldwide is continuing to grow each year. This paper presents the shift towards free cooling in data centers. It incorporates the classical techniques of case based reasoning and decision trees to be used in the decision-making process of free cooling. As snapshots of our analysis, suggestions appear in the following equations:

$$(T<68) \wedge CM \rightarrow (FC_n \wedge FC_d)$$
$$(T<68) \wedge (\neg CM) \rightarrow (FC_n \wedge \neg FC_d)$$
$$RH_m > \mu \rightarrow DH$$
$$RH_a < \mu \rightarrow (DH \leftrightarrow CM)$$

where we use predicate calculus notations. Here, T is the temperature in °F, CM denotes cooler months, FC stands for free cooling, RH is relative humidity, DH represents dehumidifiers used, m and a are subscripts for *morning* and *afternoon* respectively, while n and d are subscripts for *night* and *day* respectively. Details on these equations will be explained in our data analysis section in the paper.

This work would be of interest to data center operators, environmental management professionals, researchers from data mining and decision support communities, and anyone working in the sustainability area.

II. Background on Sustainability

The three pillars of sustainable development are based on economics, environment and society. From a broad perspective it makes sense to begin with a societal level approach in that there is a demand by society to reduce the amount of electricity, and the resulting carbon dioxide produced from running data centers. In addition, servers that make up a data center have relatively short operational lives, and there is also a societal demand to reduce or reuse electronic waste. Policymakers, business leaders, academia, environmentalists and the general public are increasingly demanding what has been called a ‘cradle to cradle’ approach to designing and producing manufactured goods. This ‘cradle to cradle’ approach involves taking a long term view of the product from where the initial materials are derived to the end use or recycling of the product. Currently, while there are numerous laws on the books towards electronic waste, there is a strong societal need for a sustainable solution.

Economics has been the primary driver in data center management to reduce the total cost of ownership. The total cost of ownership views the costs associated with purchasing and operating a server over the life time of the product. Most servers in data centers are replaced every three to four years, and the cost to run a server over that

978-1-4244-8551-2/10/$26.00 ©2010 IEEE
and ICIAfS10

time period is increasing. In many cases, the operating costs are more than the initial purchase of the original server. Due to this cost structure, recently in the USA, the Environmental Protection Agency (EPA) established Energy Star guidelines for servers [1]. This environmental standard will make the complex decision-making process of building a data center a bit easier, since the data center manager can limit the selection of servers to Energy Star products. However, as more server companies fill the market place with Energy Star servers, the data center manager will be faced with the original problem of the overwhelming selection of numerous decisions with the broad goal of attaining sustainability.

When faced with such a challenge, we claim that a decision support system is highly useful. More specifically, we suggest that case based reasoning (CBR) and decision trees are of particular value in designing such a system. Before giving further details on CBR and decision trees in this context of this problem, we now give a summary of our data analysis, followed by some details of the free cooling approach with regards to data centers.

III. Data Analysis

The purpose of data analysis for free cooling is to examine the historic temperature and humidity ranges for our study at Montclair State University (MSU) in New Jersey, approximately fourteen miles west of New York City. We focus on two parameters for our analysis in this paper, i.e., temperature and humidity as described below.

Temperature: This parameter represents the outside temperature in degrees Fahrenheit. Figure 1 presents a chart of the high and low average monthly temperature over the year at MSU where the x-axis depicts the month, and the y-axis depicts temperature in °F. The green line shows the thermostat setting at 68 °F in our main data center at MSU. The purpose of Figure 1 is to show the relationship of outside temperature in respect to the inside thermostat setting, and we notice the following.

- Shown in blue, the average monthly high is below the thermostat setting of 68°F approximately six months out of the year. Thus, free cooling could be implemented during the cooler months.
- Shown in red, the average monthly low is below the thermostat setting of 68°F, and this suggests that even in warmer (i.e. not cooler) months, free cooling could be used at night, but not during the day.

From the observations on the chart in Figure 1, our suggestion is that in the cooler months the data center could utilize free cooling instead of running the air-conditioning system throughout the year. Hence, we empirically obtain the following equations. Consider that T is the temperature in °F, CM represents cooler months, FC is free cooling, while n and d are subscripts for *night* and *day* respectively,. Thus, we get equations (1) and (2) below.

$$(T<68) \wedge CM \rightarrow (FC_n \wedge FC_d) \text{ ---- } (1)$$

and $$(T<68) \wedge (\neg CM) \rightarrow (FC_n \wedge \neg FC_d) \text{ ---- } (2)$$

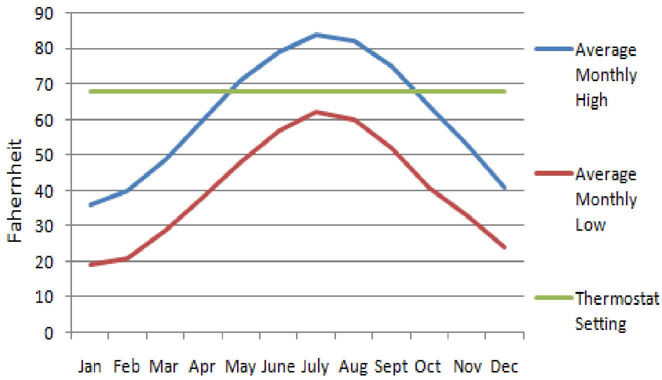

Fig. 1. Temperature Ranges at Montclair State University

Humidity: Free cooling is not without obstacles. One main challenge in applying free cooling to the local climate is that relative humidity levels need to be taken into account. Relative humidity is the amount of water vapor in a given amount of air, and is measured as a percentage. Figure 2 is a plot of humidity (expressed as a percentage) on the y-axis recorded over each month on the x-axis. Thus, Figure 2 reflects the fact that in New Jersey, relative humidity levels are high over the course of the year for free cooling. In order to implement free cooling in the local climate, we suggest that dehumidifiers would need to be installed for the incoming air to be within the recommended threshold level of 40-55% by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). While based in the USA, the focus of ASHRAE is to advance technology and promote a more sustainable world. ASHRAE does not recommend running a data center outside the desired humidity threshold, more experimental and innovative companies such as Yahoo, Intel, and Google have pushed these threshold levels in some of their data centers. In Figure 2, we notice the following;

- Shown in purple, the morning relative humidity level is clearly above the maximum threshold level of the ASHRAE high (55%). This reflects the fact that natural relative humidity levels are higher in the evening.

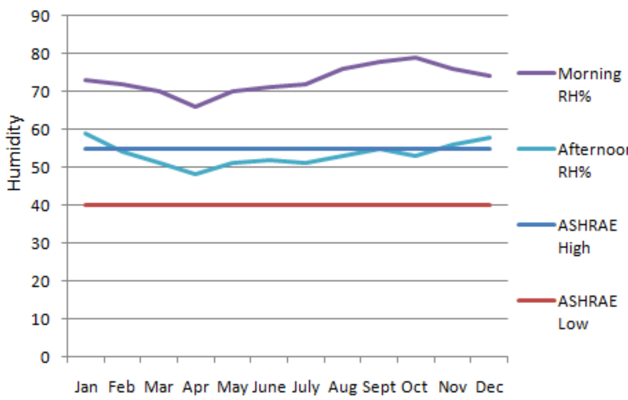

Fig. 2. Relative Humidity Levels in Northeastern New Jersey

- Shown in light blue, the afternoon relative humidity level is generally below the maximum threshold level of the ASHRAE high, except for December and January.

Hence, we now arrive at the following equations. If *RH* denotes relative humidity, *DH* represents the use of dehumidifiers, μ=threshold and the subscript *m* denotes *morning*, we get equation (3) below.

$$RH_m > \mu \rightarrow DH \text{ ---- } (3)$$

Using the same notation with subscript *a* for *afternoon*, and *CM* denoting cooler months we get equation (4) below.

$$RH_a < \mu \rightarrow (DH \leftrightarrow CM) \text{ ---- } (4)$$

The presented equations are heuristics, because the data analysis is experimental. These would be applicable in specific situations where similar observations on temperature and humidity are encountered. In the next section, the concept of free cooling is further explained from a general perspective.

IV. FREE COOLING APPROACH

We have found that the current trend towards free cooling is accompanied by a shift of locating data centers close to lower electrical costs due to the great power demands of data centers. Two areas in the USA that have experienced new construction of data centers are the Columbia River area of Washington and Oregon, and increasingly, the Buffalo region of New York. Both these areas have access to cheap hydro-electric power, and the data centers are located close to the power generation source to prevent loss of power in transportation. Hydro-power is generally viewed as a sustainable power source since it does not emit carbon dioxide; however, the construction of new hydro-power is a very controversial issue among environmentalists.

The other important variable in the establishment of data centers in the Columbia River area, and western New York State is the climate. Both areas have climates that are cool on an average basis, and allow for data centers that do not have to rely on cooling equipment year round. Recently, Yahoo began operation in 2010 of a data center in the Buffalo region of New York State that does not have conventional air cooling capabilities, and instead relies on a flat construction design that has been modeled after a "chicken-coop" architecture, in that the buildings were built out instead of up to allow for dispersion of heat. The site takes advantage of the prevailing winds of Lake Erie that are incorporated in the design of the data center complex.

Humidity levels are a concern for data center operators with the conventional wisdom that very low humidity can cause static electricity problems, and very high humidity can cause moisture problems developing on equipment. ASHRAE recommends a relative humidity range of 40% to 55% [3]. There has been a recent trend by data center operators to relax or even eliminate humidity controls [4].

As some data center operators experiment with humidity levels, another aspect of free cooling is the degree of air filtration of incoming air in free cooling data centers. Filtration of incoming air is a concern in free cooling data centers because there is a concern of limiting the amount of dust, pollen and other particular matter entering the data center. Proper air filtration and routine maintenance can solve this potential problem.

The Power Usage Effectiveness or PUE is an industry ratio for measuring efficiency in energy usage in data centers. The PUE equals the amount of total facility power used by the data center divided by the power used for the IT equipment as shown in equation (5) below.

PUE =Total Facility Power / IT Equipment Power --- (5).

The objective of data center operators is to strive for a PUE of 1.0 since theoretically that would represent a data center that is ultra energy efficient. For example, a data center that uses 125,000 kW of total facility power and 75,000 kW used for IT equipment power would have a PUE of 1.67. Standard data centers have typical PUEs of around 1.7 to 2.0, and from industry standards a better managed data center have PUEs of 1.4 to 1.6 [6],[7]. Given this general description of free cooling as a paradigm shift in data centers, we now explain decision support in the free cooling area.

V. DECISION SUPPORT

A decision support system is a tool designed to assist the users' decision-making process in a given application. Such systems are used throughout the world today in fields such as medicine, management, and other areas where complex organization is needed. With specific reference to data centers there are issues such as temperature, humidity, performance factors, and operating costs. Hence, there is a need to assist decision-making on whether and when free cooling should be used in data centers, and also on the extent of free cooling. We thus use the two decision support techniques of case-based reasoning and decision trees.

Case-based reasoning (CBR) for free cooling: CBR is structured on the concept of solving a new problem based on past experiences by reapplying information and knowledge from previous cases [8]. In our work, CBR can be viewed as a higher level technique towards building a case library of data center metrics and operations. One important goal of our project in data centers is to build a library of case studies that can be retrieved when a new data center is being designed or retrofitted. While all data centers cannot be free cooled 100% of the time, depending on the geographic location, free cooling can be incorporated into the design to minimize air-conditioning. For example, if the data center is located in the higher latitudes of the world, it can use free cooling during the cooler months. Figure 3 is an example of applying CBR to a free cooling approach in data centers. The CBR model used here is based on the classical R4 cycle of retrieve, reuse, revise, and retain [8], explained with specific reference to our problem as follows.

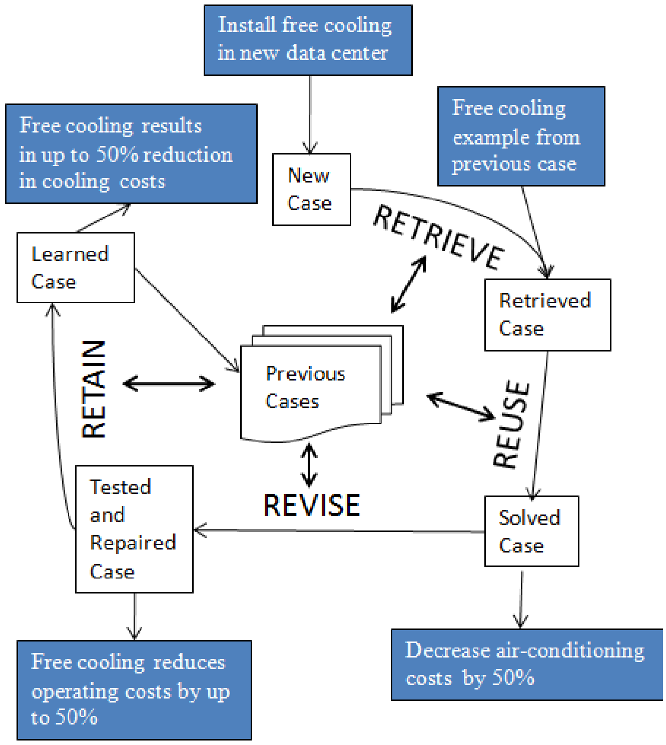

Fig. 3. CBR R4 Cycle with regards to a Free Cooling Case

- **Retrieve:** In this context of designing a new data center, the focus would be to retrieve examples from previous cases. The design team would need to examine previous cases of data centers to determine whether free cooling could be used as a cost savings measure. Important considerations would be the latitude of the proposed site, and the local weather patterns. The design team would need to retrieve such critical information as temperature and humidity levels throughout the year to estimate the number of days free cooling could be used, and if the added costs for air-circulating ventilation equipment justify the investment.

- **Reuse:** An important aspect in the design stages of the new data center is the ability to reuse previous cases. A current problem is that there are a limited number of developed cases on free cooling because originally data centers were not designed with energy efficiency as a top priority. It has only been in the last few years that data center design has evolved to examine the total cost of ownership, and seek ways to reduce operational costs. The original design specifications for data centers were to maintain a steady temperature of around 68°F, and a relative humidity between 40 and 55%.

- **Revise:** Once previous cases are retrieved, it is important to carry out *adaption* in CBR [8] to fit the current situation. In data centers, for example, there are extreme variations in temperature between the cold or air intake aisles, and the hot or exhaust aisles. While 100% free cooling has been used in some applications, it may be too risky a scenario for most organizations. Free cooling can reduce air conditioning costs by up to 50% by installing added circulating and ventilation equipment depending on location. Instead of taking hot exhaust air and cooling the air back to 68°F at a great energy expense, data centers should be designed to exhaust hot air to the outside or use that hot air for another purpose. The recycling of waste heat has been studied in industrial ecology. Likewise, as shown in this example, it is important to revise existing cases based on general domain knowledge in environmental management, as well as case specific knowledge on the given data center.

- **Retain:** The final step in the CBR model is to retain previous cases, and build a library of data centers that have applied free cooling. Building such a library presents a challenge since each data center is unique in requirements. However, as more organizations move towards free cooling, the ability to discover knowledge from previous cases is critical for sustainability.

Decision trees for free cooling: The second technique for decision support that we propose to use in this paper is decision trees. As widely known in data mining, 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 [9]. Presented in Figure 4 is a diagram of a decision tree representing three options in designing or retrofitting a data center; 1) 100% free cooling, 2) 50% free cooling, and 3) the traditional method. We expand on these options below.

1) **100% free cooling**: While some limited protocol data centers have been built using 100% free cooling [4], the main disadvantage is that there is the risk of the unknown in that the data center could fail primarily due to excessive heat if there is no back up cooling system in place. Another concern is equipment failure due to humidity exposure, but humidity controls could be added at an extra cost to maintain ASHRAE humidity standards. The advantages of a 100% free cooled data center are the savings in air-conditioning equipment, and monthly electrical bills for operating air-conditioners.

2) **50% free cooling:** This represents the concept of maximizing outside air for cooling during the seasonal variations of the year. For example, in the northeast of the USA, the 50% free cooling approach would have installed air-conditioning to be used in the summer. During the cooler months, the system would employ free cooling, and thus have the advantage of savings on monthly electrical costs for air-conditioning. The upfront design would be more complex, especially if an

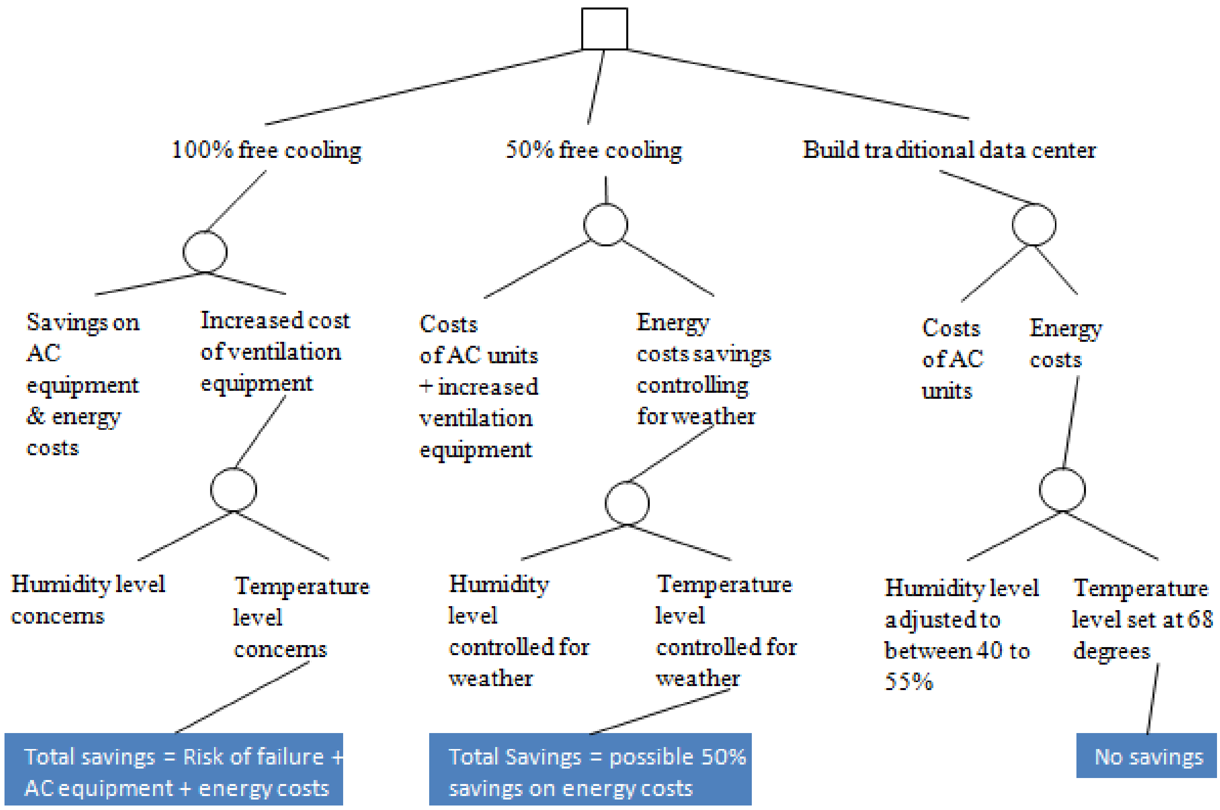

Fig. 4. A Decision Tree Example on Free Cooling Alternatives

industrial ecology framework was employed, and an added cost would be for the ventilation equipment. Conversely, when viewed from the perspective of the total cost of operation, the long term savings on the monthly electrical bill, and any savings from employing an industrial ecology framework have the ability for economic savings. They also reduce the carbon footprint of data center operations. This strategy of free cooling allows for greater variation in both temperature and humidity levels that ASHRAE expanded in 2008 [3].

3) **Traditional method:** This involves using air conditioning throughout the year as done conventionally. While this method presents the most risk adverse leaf in the decision tree, it is argued in this paper that data centers need to further evolve and push the limits of innovation. Data centers represent a growing 2% of total worldwide energy usage, and the resulting carbon footprint [9]. As data center operators shift their view to a long term approach to building data centers of the future, the total cost of ownership will be the economic driver towards savings.

VI. Further Challenges

With this discussion on the emphasis of CBR and decision trees for decision support in free cooling, we now address further challenges and proposed solutions.

1. **High demand on data centers:** Data center customers expect service 24 hours per day, 7 days a week with 100% operational time. Due to these high expeditions, data center operators are prone to be risk adverse. Therefore, a psychological hurdle is that new technologies, such as free cooling, need to be tested and approved by ASHRAE. We propose that important considerations towards change in data centers are building codes based on recommendations of groups such as ASHRAE. In order to encourage the free cooling approach and to incorporate industrial ecology, we suggest that local building codes need to be flexible to emerging technologies.

2. **Planning ahead:** Consider the following example. The current work at MSU has been to monitor and calculate the cost of running a traditional data center at a mid-to-large scale university in the northeast of the USA. The main data center is located on the 6^{th} floor of a seven story building called University Hall constructed in 2005. Currently, there are no sub-meters on the air-conditioning system, so the cost has to be estimated by working with the head electrician. The electrical cost of the air-conditioning system in the data center has been estimated for the 2009/2010 school year as follows; 1,524,240 kWh x .14 cents/kWh of electricity = $ 213,394. Therefore, had the MSU data center been designed to incorporate a 50% free cooling approach, we would estimate that MSU could be saving over $100,000 per year in air-conditioning costs, minus the investment for the initial ventilation system. In the future, we suggest that before the next air-conditioning system is installed, that an upgrade consisting of a 50%

free cooling approach be considered. Likewise, other alternatives can be employed in institutions worldwide.

3. **Life span of air-conditioning equipment:** Air-conditioning systems have a 15 to 20 year life span. Since the data center is relatively new, there are no current plans to retrofit the MSU data center. A retrofit could be cost prohibitive since the data center is located in the middle of the floor space with no outside wall. It is important to have an outside wall to install intake and outtake ventilation equipment. We claim that this problem could be overcome by installing additional duct work. Alternatively, an industrial ecology method could be employed to heat the seventh floor. It would have been beneficial to plan free cooling into the initial design of the building. However, data centers were not built with a sustainable development mindset, due to which future adjustments need to be made.

There are various ways of implementing real-world solutions to such challenging problems. For example, currently at MSU a smaller data center in College Hall is being considered for a retrofit. The smaller data center is located in a room with an outside wall where free cooling could be used for the colder periods in the New Jersey weather. The planning is in the initial stages, and the use of free cooling is being recommended by the authors. Analogous solutions can be applied elsewhere.

VII. CONCLUSIONS

In this paper the concept of free cooling is presented to cool data centers using two decision support methods. One method presented is the use of case based reasoning to build a library of previous cases in data center design. We recommend this development of previous cases in order to advance knowledge, promote innovation, and to advance building codes in data centers. Another decision support method presented is the use of decision trees to map out the complexities when deciding on criteria of a data center. An example high level three scenario decision tree is presented here, and with suitable recommendations. The paper also outlines the challenges and proposed solutions for decision support in data centers.

The main contributions of this work include;

- Recommending a free cooling approach in data centers, especially in academic settings
- Emphasizing the need for decision support in the area of free cooling
- Proposing a framework based on CBR and decision trees for decision support in this area
- Presenting the challenges involved in this effort with proposed solutions, using real-world examples
- Bridging the areas of computer science and environmental management in an interdisciplinary research activity
- Heading towards a paradigm shift in data center design with the broader goal of sustainability, an important mission across the globe today.

ACKNOWLEDGMENTS

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. 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 MSU. We also thank Dr. Dibyendu Sarkar, head of the Ph.D. program in Environmental Management for his support and cooperation. We express our gratitude to the graduate Dean at MSU, Dr. Constantine Theodosiou, for his encouragement and support. Finally, we thank the Dean of the College of Science and Mathematics, Dr. Robert Prezant, for initiating this activity, giving valuable feedback at various stages, and providing funding.

REFERENCES

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

[2] Intel Information Technology (2008), "Reducing Data Center Cost with an Air Economizer," [White Paper] Available: http://www.intel.com/it/pdf/Reducing_Data_Center_Cost_with_an_Air_Economizer.pdf

[3] ASHRAE (2008) "Environmental Guidelines for Datacom Equipment-Expanding the Recommended Environmental Envelope," Available: http://tc99.ashraetcs.org/documents/ASHRAE_Extended_Environmental_Envelope_Final_Aug_1_2008.pdf

[4] J. Stein, "Waterside Economizing in Data Centers: Design and Control Considerations," ASHRAE Transactions, 2009, Vol. 115, part 2.

[5] C. Belady, A., Rawson, J. Pfleuger, and T., Cader, "Green Grid Data Center Power Efficiency Metrics: PUE and DCIE," The Green Grid, 2008, [White Paper].

[6] ACM News, "Grants Support Design of Ultra-efficient 'Green' Data Center," Communications of the ACM, June 2010.

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

[8] A. Aamodt & E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," AICom-Artificial Intelligence Communications, IOS Press, Vol. 7:1, pp.39-59, 1994.

[9] J. R. Quinlan, "Induction of Decision Trees," Machine Learning 1, pp 81-106, 1986.

[10] J. Koomey, "Estimating Total Power Consumption by Servers in the U.S. and the World." Stanford University, 2007.