A Practical Method to Reduce HVAC Energy in Data Centers

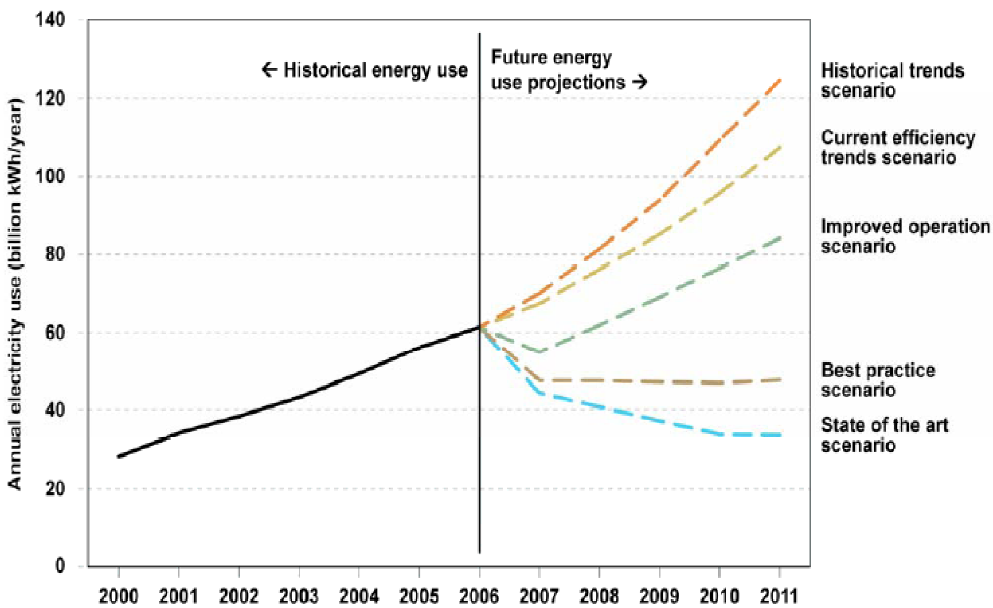
William Dunckel and Francois Xavier Rongere, Pacific Gas and Electric Company

ABSTRACT

Data Centers present a good opportunity for energy efficiency. In 2006 PG&E estimated that data centers accounted for over 400MW of power and 2.5Twh/yr of energy in PG&E service territory. Since then several new data centers have been built and existing data centers expanded in Northern California and there is no indication that this will trend will end soon.

Over the last decade many data centers were built with the focus on creating new capacity quickly rather than on cost of operation. Consequently, many exist today that are inherently inefficient from the standpoint of energy use. The bright side is that this creates the potential to save a large percentage of their energy use through energy efficiency efforts.

Today, there is a new market situation for data centers. The quest for increased computing capacity from the device to the facility level continues. However, data center operators today are often looking for ways to cut costs and Information Technology directors are looking for ways to comply with corporate green initiatives. This is an excellent setting for offering retrofit measures to help reduce data center electricity use and the associated environmental foot print.



Source: EPA report to the US Congress August 2nd, 2007

The challenge for PG&E was to develop retrofit measures that would not interfere with data center operations but would provide the short payback Silicon Valley companies expect. This paper explains how one successful measure was developed and deployed rapidly.

Introduction

The State of California has chosen customer energy efficiency as the most desirable strategy to meet a constantly growing demand for energy. In response, PG&E has discovered new ways to save energy in areas previously overlooked, specifically data centers.

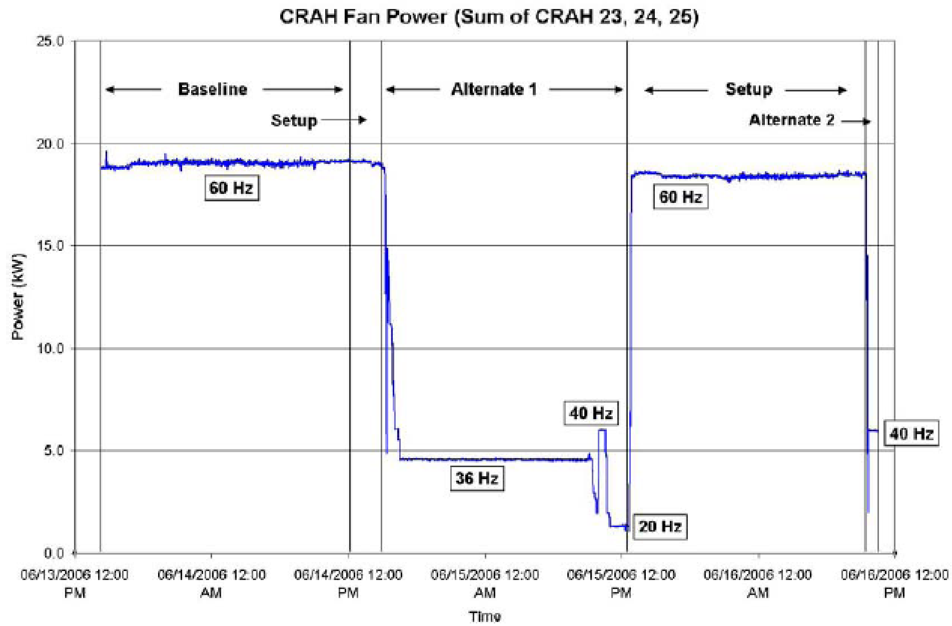
A few things came together to make PG&E the leading utility in data center energy efficiency. Silicon Valley provided a culture of innovation and technology and customers eager to try new ideas. The dot com boom and bust left behind many data centers built quickly but not necessarily for efficient operation. Many large high tech companies in the PG&E territory were experienced in energy efficiency and PG&E programs. This combination enabled PG&E quickly develop experience and expertise in data center energy efficiency.

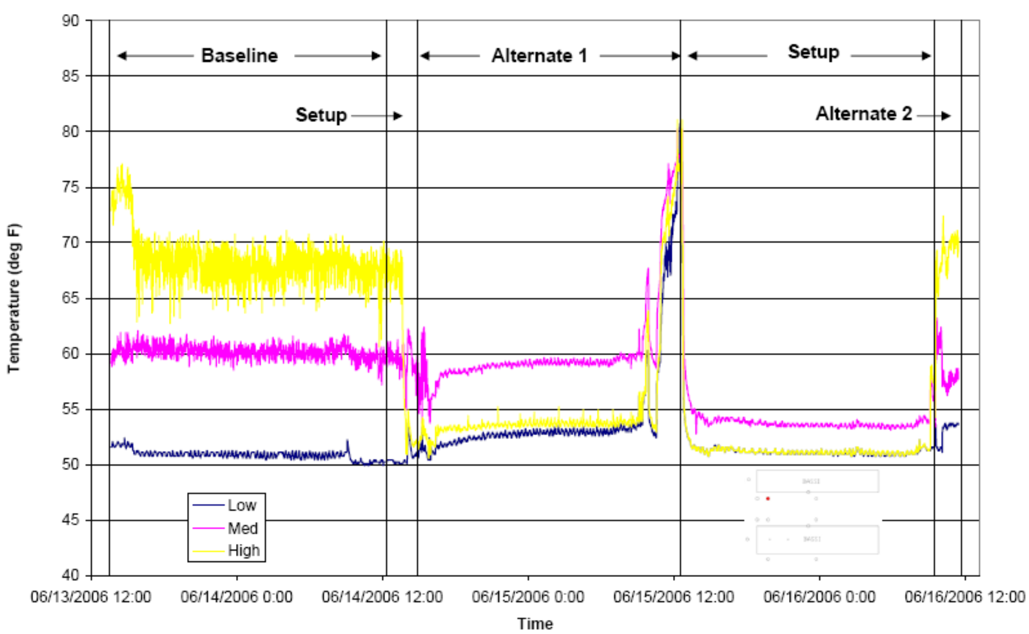
Data centers account for less than 5% of the energy consumed by PG&E customers but a single data center can consume more energy than a commercial building ten times its size. Further, it is reasonable to save a significant portion of the energy in a typical data center. The data center energy efficiency projects PG&E has completed to date have produced excellent savings per square foot compared to projects in other types of facilities.

While data centers became prime targets for PG&E's CEE outreach efforts there were many hurdles to overcome. The biggest was the fear of making change that could disrupt operations or reduce reliability. An approach was needed that would address these concerns while providing opportunities to save within a short payback period.

In 2005 PG&E and LBNL conducted an experiment to test the effects of isolating the cold supply air stream from the rest of the data center. Two cold aisles were fully isolated from adjacent hot aisles and cooling unit fan speed was adjusted while server inlet temperatures were monitored. The figure below shows how fan speed was reduced (Alternate 1) from the baseline of 60 Hz to 36 Hz at which point further reduction in fan speed caused the server fans to speed up. The next figure shows the server inlet temperature at the top, middle, and bottom of the racks. Before isolation, the baseline showed temperatures ranged from about 52 to 68F. With isolated aisles (Alternate 1), the temperatures ranged only from about 53 to 60F. This demonstrates that significant fan savings can be accomplished while also improving the uniformity of temperature distribution to servers.

PG&E was approached by a customer in search of ways to increase cooling of a small but dense data center. A visit to the site revealed that hot and cold air streams were mixing and the existing cooling units had more then enough capacity to serve the heat load. PG&E proposed to work with this customer to isolate air flows with the expectation of meeting cooling needs with the existing HVAC equipment. This proved to be successful and became the pilot of what is now called PG&E's Air Flow Upgrade Incentive.


CRAH Fan Power (Sum of CRAH 23, 24, 25)
Baseline
Setup
Alternate 1
Setup
Alternate 2
60 Hz
36 Hz
40 Hz
20 Hz
60 Hz
40 Hz
Power (kW)
25.0
20.0
15.0
10.0
5.0
0.0
06/13/2006 12:00 PM
06/14/2006 12:00 AM
06/14/2006 12:00 PM
06/15/2006 12:00 AM
06/15/2006 12:00 PM
06/16/2006 12:00 AM
06/16/2006 12:00 PM
Time


Baseline
Alternate 1
Setup
Setup
Alternate 2
Temperature (deg F)
90
85
80
75
70
65
60
55
50
45
40
Low
Med
High
06/13/2006 12:00
06/14/2006 0:00
06/14/2006 12:00
06/15/2006 0:00
06/15/2006 12:00
06/16/2006 0:00
06/16/2006 12:00
Time

PG&E's Air Flow Upgrade Incentive is a special method to retrofit data centers. It is based on measuring the difference between supply and return air temperature of the cooling units compared to the 28F difference from the front to the back of a typical server. If the hot and cold air paths are completely separated and the cooling CFM is perfectly matched to the total compute equipment CFM, the temperature across the cooling units can be raised to 28F rather than the typical 12-14F. Since most data centers are over provisioned with cooling units, it is often possible to turn off some cooling units and allow the others to handle the load at higher delta T. The savings comes from taking the excess cooling units off line. Alternately, if the cooling units are equipped with variable frequency drive fans, savings can be accomplish by adjusting cooling unit fan speeds.

PG&E has completed several projects with favorable results. In each case customers are saving HVAC energy while also decreasing the variance in server inlet temperature significantly. In addition, paybacks have been sufficiently attractive to convince customers to implement. Depending on the existing conditions and HVAC system, paybacks have ranged from 6 months to 3 years. Longest paybacks have been associated with converting cooling units from fixed to variable air volume.

Literature Review

Many data centers were built or expanded during the "dot com" boom with the primary goal of providing more computing capacity quickly. In addition, power density of computing equipment continued to increase. When hot spots developed facilities personnel did not have much information on how to best cool these hot spots. As noted by Karki et al in 2003, "However, because modifications in one region of the floor influence flow rates throughout the floor, considerable trial and error is involved in identifying adjustments that will yield the desired changes in the flow rate distribution. This design practice is time consuming and expensive, and often the resulting arrangement is not optimum." Karki et al proposed and tested the use of computational fluid dynamics modeling as a tool to aid in developing the optimal HVAC adjustments. [1]

Further attention was given to dealing with the complexity of data center cooling that confirmed how the common intuitive approach of adjusting floor tiles and other HVAC measures is not adequate. Proper rack layout combined with capacity provisioning of CRAC units is critical. [5].

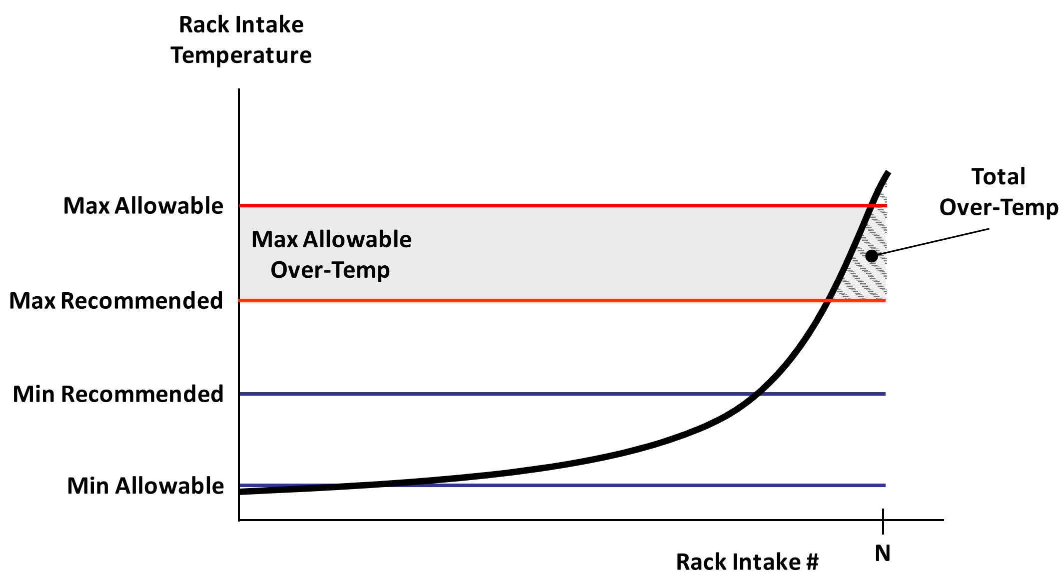
Given the continued interest in employing CFD modeling as the primary tool to guide proper adjustment of HVAC in data centers, PG&E briefly investigated the potential of offering this service to customers as a means to encourage more energy efficient data center cooling. While CFD modeling is an excellent tool to predict the impact of facility modifications PG&E would still have to find a way to calculate the potential energy savings of modifications the customer would elect to implement. PG&E continued to look for other methods to convince data centers to improve energy efficiency.

A good deal of attention had been directed to modeling air flow both under and above the raised floor. Early on it was understood that uniform airflow rates through floor tiles was the desired target and the parameters behind this were investigated. [2] A number of practical techniques to control air flow distribution were studied. One study was published that proposed installing thin partitions under the floor to direct air flow to help make under floor pressure more uniform. [3] While this can be a method to improve HVAC effectiveness and potentially improve HVAC efficiency, it is difficult to show that the solution will continue to provide the desired results for very long. As mentioned in most literature on CFD modeling for data center cooling, the heat load on the racks is generally neither evenly distributed nor static. As the hot spots move, the HVAC solution needs to respond. Under-floor partitions would have to be adjusted as often as economically possible and that would require the expense of periodic CFD models and associated measurements. This was not the solution needed for PG&E's energy efficiency programs which require that energy savings will persist for several years.

Credit should be given to those leaders who use CFD modeling and experimented with HVAC adjustments to demonstrate some fundamental concepts in data center cooling including the importance of uniform under-floor pressure and the potential to improve HVAC efficiency through more effective air flow. They showed the industry how intuitive adjustments of the HVAC system are not adequate to optimize HVAC efficiency. This also helped prompt others to search for practical solutions that could be implemented by the smaller data centers unable to venture into the world of CFD modeling.

By 2006 a number of HVAC solutions had been implemented and studied. William Tschudi et al. [4] noted that a useful metric if HVAC effectiveness and efficiency is the ratio of HVAC to IT energy. They further notes that this ratio varies five fold between the best and the worst of the data centers they studied. In addition to listing a number of practical measures for improving HVAC, they noted that air management is a key factor and is focused around isolating air flow in hot aisles from that of cold aisles.

It is difficult to know where to focus first to improve the efficiency of a complex system like data center HVAC. Fortunately, The Uptime Institute studied this matter and found that on average "2.6 times more cooling capacity operating than critical load in (the) data center" and that the 'greatest offender is Bypass Airflow – only 40% of the cold air goes directly to cool the critical load" (Sullivan 2007) [7]. Although most small to medium data centers utilize constant air volume HVAC systems, this information implies that it may be possible to shut off some of those units if bypass air is minimized and under-floor pressure can be kept relatively uniform. That could be the sort of solution PG&E could suggest to data centers to save energy.



$$RCI = 1 - \frac{Total\ Overtemperature}{Max\ Allowable\ Overtemperature} * 100\%$$

With the advent of aisle isolation solutions came the need to measure the results so data center operators can be confident that their time and money is well spent and that they are meeting the cooling requirements of the rack equipment. ANCIS Inc. [6] developed the rack cooling index show above based on the range of inlet temperature the data center is specified to achieve. In order to optimize HVAC performance, a metric like this is needed to first determine how well the HVAC system is serving the cooling needs of the rack equipment. Such a metric should be integrated into studies of proposed solutions to ensure HVAC efficiency does not hard IT operations. This is a key element PG&E needed to confidently promote air flow management improvement concepts to customers.

Methodology

The 2003 Rocky Mountain Institute Charette identified several areas for energy efficiency improvement. Two of them, air side economizers and air flow improvement, were chosen by PG&E as potential areas for further study as potential program offerings. Two projects were completed by PG&E's Emerging Technologies group to test the effectiveness of aisle isolation as a means to improve air flow management.

A preliminary air flow model as developed in Excel to determine the amount of excess cooling installed in a data center. The main idea was to compare the heat generated by rack equipment with the cooling and ventilation capacity of the cooling room units. First, the rack kW needed to be converted to required cooling CFM. Next, the difference in temperature across the rack needed to be established. These two numbers, CFM and rack delta T, then determined the required cooling capacity. By comparing that to the capacity of the installed cooling units, the excess cooling capacity could be determined.

Shortly after this model was being formed, an opportunity presented itself at PG&E customer site. This customer contacted PG&E to discuss their challenges in cooling their 2,000 sqft corporate data center with about 150w/sqft of rack load.. They were concerned that it took most of a weekend to bring their data center back to target temperature after an HVAC outage. A visit to the data center revealed that this customer had the typical flooded supply and flooded return layout with very little to segregate hot from cold air flows. Large gaps existed in racks allowing air to flow right through. Nothing was in place to prevent air from flowing unobstructed over the tops and around the ends of the racks. Furthermore, a temporary cooling unit was in place to blow cold air on one rack and exhaust hot air back to the room cooling units. All cooling units were running constantly to maintain room temperature.

PG&E's response was to propose an experiment based on what PG&E had learned about air flow management. The customer agreed and a pilot project was launched to enclose cold aisles. This first step after discussing the concepts of air flow management was to finish arranging the data center into alternating hot and cold aisles (HACA). Fortunately, equipment was facing the wrong way on only one rack and there was enough aisle width to put a partition down the center of the aisle.

With HACA completed the next step was to place temperature sensors in the supply and return air streams of each cooling unit and in various places in the cold aisles. This equipped PG&E to determine excess cooling and equipped the customer with a means of monitoring how the physical changes they made were affecting temperatures in the room. This measure-as-you-go approach provided the customer the control they needed to feel safe about making changes in their mission-critical data center.

After collecting temperatures every 15 minutes for a week, the sensors were downloaded and the data reviewed graphically. Visual inspection revealed that temperatures varied only slightly and could be considered constant to the sake of this project. Only the average temperature of each sensor was used. The data showed that temperature difference across cooling units was 12 to 18 degrees F and the variance in cold aisle temperatures was about 14 degrees F. The PG&E model indicated that about 45 kW of excess HVAC energy was being used and could possibly be saved with improved air flow.

The customer proceeded to install blanking panels in racks, strip curtains at the ends of cold aisles, and partitions above racks to force air through the servers. At the same time they install in row coolers that pull air directly from the hot aisle, cool it, and return it to the cold aisle. At then end of the project, they were able to shut off two cooling units and remove the temporary cooling unit. Net HVAC demand was reduced close to 45 kW and cold aisle temperature variance was reduced to two degrees.

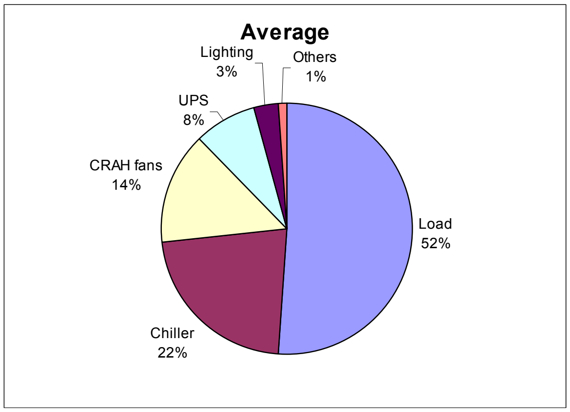
This experience encouraged PG&E to create a special air flow upgrade incentive for data centers. Most data centers are reasonably small and are interested in saving energy either to cut cost or to facilitate growth of the IT function. This incentive is a cost effective alternate to the classic energy audit that does not necessarily result in an energy efficiency project. It has also proven to be something that data center service providers can incorporate into their offerings. This enables vendors to promote data center energy efficiency to the vast number of small data centers that PG&E cannot identify.

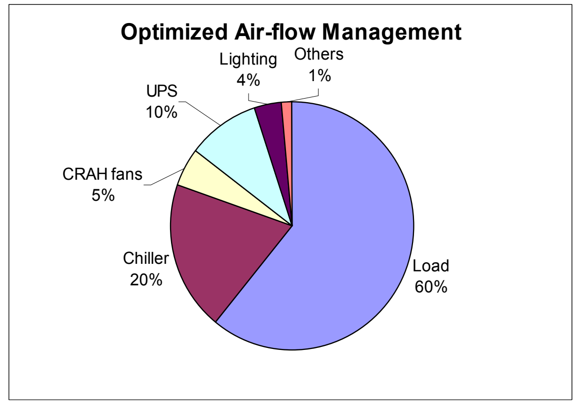
PG&E structured a special incentive specifically to capture energy savings in small data centers through air flow upgrades. This incentive is offered under the Standard Performance Contract nonresidential retrofit program that has been in place in California for many years. A special process was developed where PG&E would provide sensors and direct the customer on the placement of them. PG&E would collect the sensor data and use it to estimate the potential savings possible if the customer could implement perfect air flow. When the customer completes upgrades and shuts off CRAC units or slows down the fan speed of CRAH units, then PG&E would return to take measurements again, validate final conditions, and pay an incentive based on what was turned off or slowed down subject to a maximum of half the project cost.

Results

The air flow upgrade incentive is proving to be economical tool for PG&E to capture energy saving retrofit projects in small data centers. It avoids the complexity and time consumption of the typical measure & verify program methodology. It also is much less expensive that the engineering analysis, recommendation, and verification processed typically deployed for large retrofit and new construction projects.

Since completing the pilot air flow upgrade project in April 2008 PG&E has captured over two dozen projects. One customer was able to realize a 44.9 kW, 392.1 MWh savings of the 45 kW, 394.1 MWh potential savings identified by implementing both aisle isolation and the enhancement of the HVAC system with in-row cooling units with VFD fans. Another customer realized 78 kW, 683.7 MWh of the potential 116.9 kW, 1023.6 MWh with aisle isolation and no modification of the HVAC system. This second case is more typical of what PG&E expects to see where customers correct location of perforated tiles, install blanking panels, install strip curtains, and minimize floor openings to achieve a reasonable degree of aisle isolation. Savings will vary to the degree aisle isolation is completed and the degree to which the excess CFM matches the CFM of available cooling units that can be reduced.





The model for estimating potential savings has been refined to handle the presence of outside air economizers for one climate zone with additional climate zones to be added as needed. The basic model design remains intact and is providing acceptable estimates for the typical data center situation.

PG&E has been contacted by several firms who are interested in incorporating this incentive into the variety of services they offer their data center clients. Interest ranges from firms who mainly offer sensors and reporting systems to firms who offer turn key data center audits and operational improvements. A few projects have already been completed where vendors introduced the incentive to their clients and coordinated the interaction with PG&E.

After correcting placement of perforated floor tiles and plugging holes in the floor, strip curtains and blanking panels are popular was to isolate aisles. Note that this customer had fire suppression in all aisles allowing them to install fixed curtains and partitions above racks. Another popular solution is to use strip curtains suspended by fusible links at aisle ends and above the face of the racks.

One unexpected result arose and should be noted. This incentive has limited applicability to computer labs where the majority of rack equipment is not servers. Labs that test network gear, for example, may have a predominance of network equipment that does not raise air temperature any near the 28 degrees of the typical server. In that case, there is very little if any advantage in isolating aisles. When the delta T across the rack is less than 20 degrees F or so, it may not be possible to raise the return air temperature enough concentrate the cooling load onto a portion of the cooling units and shut off the excess cooling units. In such a facility, it may still be possible to adjust the distribution of floor tiles or rack equipment to minimize hot spots and consequently, reset the room set point a few degrees to obtain some savings. However, this will be a small portion of the amount of savings possible in a server-dominated facility and much more difficult to maintain. The customer who brought this to the attention of PG&E has since decided to try grouping all servers into one area separated from the networking gear and implementing cold aisle isolation in that server area only.

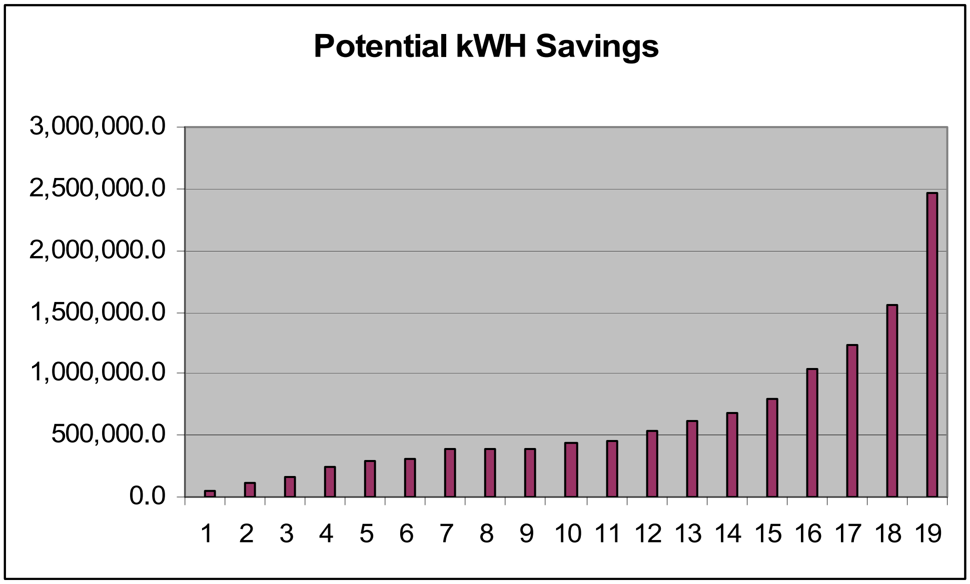
It was discovered quickly that the cost of implementing aisle containment went up significantly when the fire suppression system was involved. In order to prevent air flow over the top of racks, customers need to install something above each rack. Typically, the fire suppression heads are located above each rack rather than in each aisle to save cost. When that is the case, any partitioning above the rack can interfere with the function of the fire suppression system. Modifying the fire suppression system can be both a significant concern to the customer as well as a significant expense. Some customers found they could split the fire suppression system to provide heads in each aisle while others utilized strip curtains hung by fused links that would drop to the floor before the fire suppression system activated.

One customer brought up the question of ride through – the ability to maintain room temperature in the event of a power failure when the rack equipment continues to produce heat but the HVAC system is down. In response, ANCIS was asked to investigate published information on this issue and present a summary of findings. The conclusion was that the ride through time decreased an order of magnitude when an open architecture was changed to closed hot or cold aisles. The servers do not have the ability to pull cold air from the under-floor plenum when the HVAC system is off line. In an open architecture, the servers can recirculate room air that will mix and provide the servers with an average temperature. When there is a large space above the racks, this can provide a large thermal mass to delay the rise of room temperature. When there is little room above the racks, this is less of an issue.

Discussion

The potential to save significant energy economically in data centers has been confirmed with the success of several air flow upgrade projects. Savings from eliminating excess HVAC energy are congruent with the findings of the Uptime Institute that revealed the extent to which data centers HVAC systems are over-provisioned and the amount of cooling that actually serves the critical rack load. In all projects, the percent of potential savings actually achieved was limited only by the customer's ability to implement aisle isolation and to shut off the optimal amount of excess cooling equipment.

The chart below shows potential annual kWH savings identified for the first 19 projects in the program. Total potential savings is 1,045 kW and 12,118,183 kWH. At the time of this writing four projects have been completed and verified with actual savings ranging from 67 to 99 % of potential savings. Note that the distribution of savings by project is similar to the distribution of data center size in the market. A vast majority is small and not found in the facilities of all commercial and industrial customer market segments.



It is interesting to note that the successful implementation of aisle isolation often requires the customer to change his understanding of process HVAC. Perhaps the most important success factor is the fundamental insight the customer needs to have confidence to move ahead with implementation. Specifically, many data center operators have been treating hot spots by directing high velocity streams of cold air on to them. Temporary cooling units are often deployed specifically for this reason. Over time, as hot spots move, the operator begins to think that any long lasting solution is unlikely or will be very expensive and complicated. It is very important to turn around this thinking by staying close to the customer during implementation.

Aisle isolation works on the principle of providing the right volume of air to the whole cold aisle. Rather than waste energy creating turbulence and moving air quickly, the volume of the cooling system air is balanced with the volume of the rack equipment. This is a more sustainable solution that helps ensure that energy savings will persist. Once a data center is converted to isolated aisles, the hot spots are diminished or even eliminated and the temperature of each cold aisle can be managed as a unit rather than the temperature of each server. The customer much switch from thinking more volume of air is better to thinking that the right volume of air is best.

The current model assumes that rack load is cooled with 120 CFM per kW although this number can easily be changed if desired. This figure can be changed if there is site specific data. With the exception of a few computer labs, this assumption has proven effective in estimating savings. As servers and other rack equipment evolves there may be need to adjust this figure. With the efforts of the EPA to encourage data center energy efficiency, and the cooperation of industry, it is reasonable to expect that this will not change unexpectedly. Also, it is expected that aisle isolation will become standard practice before significant changes in electronics ventilation are experienced.

Summary

This utility-sponsored program to guide data center operators through air flow upgrades has proven to be an effective tool for energy savings in small data centers. A large portion of the potential calculated savings is captured in the typical project and with short payback periods. It is an attractive program for both data center operators and vendors of data center services. It is an effective mechanism for the utility to influence the large number of data centers that it cannot otherwise detect and reach.

This approach may not be suitable for all sizes of data centers. When only one cooling unit serves a room there is not much opportunity to reduce cooling capacity to match rack load. Savings would only be possible of the cooling unit has a variable speed fan or if the customer was seeking to add more rack load but was restricted by the cooling system.

In large data centers two factors may make other approaches more desirable. When the room is very large it may not be possible to turn off specific CRAC units and have the adjacent units handle the load and distribute cooling over long distances. With the large potential savings of a large data center at stake, it may be more fitting to bring in engineering consultants or CFD modelers to provide more comprehensive analysis of proposed changes.

One idea for further study is to see how much of the potential savings can be achieved by isolating aisles only up to the top of the rack. If the expenses related to fire suppression can be avoided it will facilitate project implementation and potentially provide excellent payback. If the CFM of the cooling system can closely match the CFM of the rack then there will be very little to cause air mixing. Most small data centers have constant volume cooling units and cannot match supply with demand CFM closely. Fortunately, some key manufacturers offer replacement products that address this now or in the immediate future.

There is also opportunity to learn more about converting the typical direct expansion cooling unit from constant to variable air volume so cooling can better modulate to match rack load. Small changes in fan speed can produce large savings in fan energy. However, most computer room cooling units do not have controls that can allow fan speeds to be reduced and still protect the cooling coils. At the time of this writing a few customers in PG&E territory are contemplating upgrading their controls and converting to variable air volume.

Co-location facilities may find it difficult to achieve good air flow management since they are not in control of configurations at the rack level. Rather than struggle with this, co-location facilities in mild climates like California may find it more practical to implement air or water side economizers to save energy. Either air flow upgrades or economizers are promising energy efficiency measures by themselves but combining them enables the economizer to provide even greater savings.

References

[1] Chandrakant D. Patel, Ratnesh Sharma, Cullen E. Bash and Abdlmonem Beitelmal. "Thermal Considerations in Cooling Large Scale High Compute Density Data Centers", Hewlett Packard.

[2] Roger R. Schmidt, Kailash C. Karki, Kanchan M. Kelkar, Amir Radmehr, Suhas V. Patankar "MEASUREMENTS AND PREDICTIONS OF THE FLOW DISTRIBUTION THROUGH PERFORATED TILES IN RAISED-FLOOR DATA CENTERS" ,

Proceedings of IPACK’01 The Pacific Rim/ASME International Electronic Packaging. Technical Conference and Exhibition, July 8–13, 2001, Kauai, Hawaii, USA

[3] Kailash C. Karki, Suhas V. Patankar, Amir Radmehr, “TECHNIQUES FOR CONTROLLING AIRFLOW DISTRIBUTION IN RAISED-FLOOR DATA CENTERS” , Proceedings of IPACK03 The Pacific Rim/ASME International Electronic Packaging. Technical Conference and Exhibition, July 6-11, 2003, Maui, Hawaii, USA

[4] William Tschudi, Evan Mills, Steve Greenberg, Peter Rumsey, “Measuring and Managing Data Center Energy Use”, HPAC Engineering March 2006

[5]Kailash C. Karki, Amir Radmehr, and Suhas V. Patankar “Use of Computational Fluid Dynamics for Calculating Flow Rates Through Perforated Tiles in Raised-Floor Data Centers”, International Journal of Heating, Ventilation, Air-Conditioning, and Refrigeration Research, Volume 9, Number 2, April 2003, pp. 153-166.

[6] Herrlin, M. K. Rack Cooling Effectiveness in Data Centers and Telecom Central Offices: The Rack Cooling Index (RCI). ASHRAE Transactions, Volume 111, Part 2, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., Atlanta, GA. 2005.

[7] Sullivan, Robert, “Legacy Data Centers -Their Problems and Solutions”, Critical Facilities Round Table Legacy Data Center Committee Meeting, August 2007