



Thermo-Fluids Provisioning of a High Performance High Density Data Center

Monem H. Beitelmal, Chandrakant D. Patel
Internet Systems and Storage Laboratory
HP Laboratories Palo Alto
HPL-2004-146(R.1)
September 24, 2004*

data center, high
density, server
failure, heat
transfer, modeling,
provisioning,
cooling

Consolidation and dense aggregation of slim compute, storage and networking hardware has resulted in high power density data centers. The high power density resulting from current and future generations of servers necessitates detailed thermo-fluids analysis to provision the cooling resources in a given data center for reliable operation. The analysis must also predict the impact on the thermo-fluid distribution due to changes in hardware configuration and building infrastructure such as a sudden failure in data center cooling resources. The objective of the analysis is to assure availability of adequate cooling resources to match the heat load, which is typically non-uniformly distributed and characterized by high-localized power density. This study presents an analysis of an example modern data center with a view of the magnitude of temperature variation and impact of a failure. Initially, static provisioning for a given distribution of heat loads and cooling resources is achieved to produce a reference state. A perturbation in reference state is introduced to simulate a very plausible scenario--failure of a room air conditioning (CRAC) unit. The transient model shows the "redlining" of inlet temperature of systems in the area that is most influenced by the failed CRAC. In this example high-density data center, the time to reach unacceptable inlet temperature is less than 80 seconds based on an example temperature set point limit of 40°C (most of today's servers would require an inlet temperature below 35°C to operate). An effective approach to resolve this issue, if there is adequate capacity, is to migrate the compute workload to other available systems within the data center to reduce the inlet temperature to the servers to an acceptable level.

Thermo-Fluids Provisioning of a High Performance High Density Data Center

Monem H. Beitelmal and Chandrakant D. Patel
Hewlett-Packard Laboratories
Palo Alto, CA 94304

Abstract

Consolidation and dense aggregation of slim compute, storage and networking hardware has resulted in high power density data centers. The high power density resulting from current and future generations of servers necessitates detailed thermo-fluids analysis to provision the cooling resources in a given data center for reliable operation. The analysis must also predict the impact on the thermo-fluid distribution due to changes in hardware configuration and building infrastructure such as a sudden failure in data center cooling resources. The objective of the analysis is to assure availability of adequate cooling resources to match the heat load, which is typically non-uniformly distributed and characterized by high-localized power density. This study presents an analysis of an example modern data center with a view of the magnitude of temperature variation and impact of a failure. Initially, static provisioning for a given distribution of heat loads and cooling resources is achieved to produce a reference state. A perturbation in reference state is introduced to simulate a very plausible scenario – failure of a room air conditioning (CRAC) unit. The transient model shows the “redlining” of inlet temperature of systems in the area that is most influenced by the failed CRAC. In this example high-density data center, the time to reach unacceptable inlet temperature is less than 80 seconds based on an example temperature set point limit of 40 °C (most of today’s servers would require an inlet temperature below 35 °C to operate). An effective approach to resolve this issue, if there is adequate capacity, is to migrate the compute workload to other available systems within the data center to reduce the inlet temperature to the servers to an acceptable level.

1. Introduction

Modern slim server designs enable immense compaction in data centers resulting in high power density – 500 W/m² today to 3000 W/m² in 5 years. The compaction enables users to consolidate multiple data centers to reduce cost of operation. However, the high power density resulting from a dense aggregation of servers results in complex thermo-fluid interaction in data centers. This requires engineering rigor in analysis and design to assure uptime required by users. In a departure from state-of-the-art approaches, -- provisioning cooling resources by capacity matching and intuition, -- future data centers must be designed with detailed thermo-fluids analysis [1]. Furthermore, once statically provisioned for a fixed distribution of resources, the data center must dynamically adapt to IT changes, such as addition of new computational loads (heat loads) and infrastructure changes such as loss of a cooling resource.

Figure 1 displays a typical state-of-art data center with modular air-conditioning units and an under-floor cool air distribution system. The computer room air conditioning (CRAC) units are sized by adding up the heat loads and the airflow is distributed qualitatively.

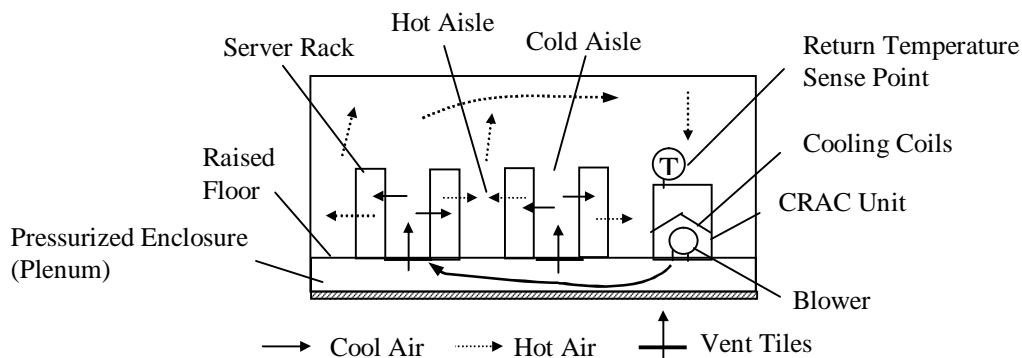


FIG. 1 Typical raised-floor data center configuration

The CRAC units are controlled by temperature sensed at the return air inlet, as shown in Fig. 1, and the “set” point is input manually based on qualitative assessment of room environmental conditions. A large data center, with hundreds of 10kW racks requires hot zones and cold zones, also called hot aisles and cold aisles. The hot return air stream must return to the CRAC units without mixing with the cool air from the supply. However, the current method of design and control results in mixing of the two streams – an inefficient phenomenon that prevents achievement of the compaction goals. Patel et. al describe this limitation and propose new approaches based on computational fluid dynamics modeling in design of data centers [1, 2]. Sharma et. al introduce new metrics to assess the magnitude of recirculation and mixing of hot and cold streams [3]. These metrics, the Supply Heat Index (SHI) and the Return Heat Index (RHI), are based on non-dimensional parameters that describe the geometry of the data center. Bash et. al show the overall design techniques to enable energy efficient design of data centers [4]. In order to adapt to changes in data centers, Patel et. al introduce dynamic provisioning of cooling resources called “smart” cooling in data centers [5]. Sharma et. al suggest thermal policies that can be used to allocate workloads [6]. Patel et. al propose global workload placement according to the efficiency index of a data center that is based on external environment and internal thermo-fluids behavior [7].

This paper discusses a method describing the provisioning of data center cooling resources based on detailed numerical modeling. This method utilizes Flovent [8], a commercially available packaged computational fluid dynamics (CFD) finite volume code. The method takes into account as its input the topological distribution of racks (heat load) and the layout of vent tiles and CRAC units (cooling resources). The results are presented in terms of the

CRAC unit capacity utilization, temperature, pressure and flow distribution. Thus, this paper serves as a case study of an example data center that has been provisioned through CFD modeling technique by examining the impact of IT and facility changes to facilitate operational procedures that enable reliable operation. The study provides details on how such analysis can be used to plan and design for these eventualities, underscoring the power of the numerical design approach. The importance of predicting the performance of a given data center prior to construction and facilitating changes to the layout of the IT and cooling resources are evident.

2. Overview of Numerical Method

High air mass flow rate coupled with the complex thermo-fluid interaction in a data center requires 3-dimensional numerical analysis. A method for provisioning data center cooling resources based on a detailed CFD modeling has been developed and validated at Hewlett-Packard Laboratories.

The dimensional governing equations of fluid flow describing the properties fields can be written as follows:

Continuity:

$$\partial r / \partial t + \nabla \cdot (r \mathbf{u}) = 0 \quad (1)$$

Conservation of Momentum:

$$r \partial \mathbf{u} / \partial t + r \mathbf{u} \cdot \nabla \mathbf{u} = -\nabla P + \nabla \cdot \mathbf{t} + (r - r_\infty) \mathbf{g} \quad (2)$$

Conservation of Energy:

$$r \partial e / \partial t + r \mathbf{u} \cdot \nabla e = \nabla \cdot (l \nabla T) - P(\nabla \cdot \mathbf{u}) \quad (3)$$

These are the fundamental equations to be solved for the flow field variables (pressure, temperature, density and velocity). Equation (1) is the continuity or conservation of mass equation, which states that the mass entering a control volume is equal to the mass leaving the control volume plus the change of mass within the same control volume. Equation (2) is the momentum equation, derived from Newton's second law and states that force is equal to mass times acceleration. In this case, the force exerted on the fluid volume could be pressure, velocity, viscous or body forces. The left-hand side of the equation is the density (mass per unit volume) times the acceleration, and the acceleration is the change of velocity with respect to time and distance in all directions. On the right-hand side of the equation are the pressure forces, the viscous forces and the gravity or body force. The latter becomes important when buoyancy is a factor. Momentum is a vector equation and therefore has magnitude and direction. Equation (3) is the energy equation derived from the first law of thermodynamics, which states that the net heat transfer into a control volume plus the net work done on the same control volume is equal to the difference between the final and the initial energy state of the same control volume.

For a three-dimensional incompressible flow and in the limit of infinitesimally small control volume, the governing equations, equations 1 thru 3, are reduced to the following equations:

Continuity:

$$\left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z} \right) = 0 \quad (4)$$

Momentum:

$$\begin{aligned}
 r \left(\frac{\partial u_x}{\partial t} + \frac{\partial u_x}{\partial x} u_x + \frac{\partial u_x}{\partial y} u_y + \frac{\partial u_x}{\partial z} u_z \right) &= \left(\frac{\partial s_x}{\partial x} + \frac{\partial t_{xy}}{\partial y} + \frac{\partial t_{xz}}{\partial z} \right) + f_x \\
 r \left(\frac{\partial u_y}{\partial t} + \frac{\partial u_y}{\partial x} u_x + \frac{\partial u_y}{\partial y} u_y + \frac{\partial u_y}{\partial z} u_z \right) &= \left(\frac{\partial s_y}{\partial y} + \frac{\partial t_{yx}}{\partial x} + \frac{\partial t_{yz}}{\partial z} \right) + f_y \\
 r \left(\frac{\partial u_z}{\partial t} + \frac{\partial u_z}{\partial x} u_x + \frac{\partial u_z}{\partial y} u_y + \frac{\partial u_z}{\partial z} u_z \right) &= \left(\frac{\partial s_z}{\partial z} + \frac{\partial t_{zy}}{\partial y} + \frac{\partial t_{zx}}{\partial x} \right) + f_z
 \end{aligned} \tag{5}$$

Where

$$s_i = -p + 2m \frac{\partial u_i}{\partial i} \tag{6}$$

$$t_{ij} = m \left(\frac{\partial u_i}{\partial j} + \frac{\partial u_j}{\partial i} \right) \tag{7}$$

Energy

$$\begin{aligned}
 r \frac{DE}{DT} = -\text{div}(p\mathbf{u}) &+ \left(\frac{\partial(\mathbf{u}t_{xx})}{\partial x} + \frac{\partial(\mathbf{u}t_{zy})}{\partial y} + \frac{\partial(\mathbf{u}t_{xz})}{\partial z} + \frac{\partial(\mathbf{v}t_{xy})}{\partial x} + \frac{\partial(\mathbf{v}t_{yy})}{\partial y} + \frac{\partial(\mathbf{v}t_{yz})}{\partial z} + \frac{\partial(\mathbf{w}t_{xz})}{\partial x} \right. \\
 &+ \left. \frac{\partial(\mathbf{w}t_{yz})}{\partial y} + \frac{\partial(\mathbf{w}t_{zz})}{\partial z} \right) + \text{div}(k\text{grad}T) + S_E
 \end{aligned} \tag{8}$$

The partial differential equations describing the complex interactions between the fluid flow and heat transfer are highly coupled and non-linear. Numerical modeling techniques (finite volume, finite element and finite difference) called computational fluid dynamics (CFD) are available to solve for the physics described by these equations. Discretization technique is used to transform the partial differential equations into algebraic form and then solved iteratively until a suitable convergence level is reached. The convergence level is determined

based on the variables (pressure, temperature, velocity) residual monitored relative to the problem scale that is being solved.

3. Modeling Case Study

The cooling capacity needed for the computational load (heat load) processed by compute, networking and storage resources in the data center is available through manufacturer specification. However, the distribution of the heat load relative to the cooling resources within the data center is essential, as it is critical to deliver sufficient amount of cooled air at a specific temperature to the inlets of the racks. Along these lines, a model of the data center requires information on the actual and maximum heat load and the total air flow and capacity of the CRAC units. The CRAC units total capacity should be adequate for the actual heat load at any given time. The model, once converged, solves for the distribution of temperature and airflow, and enables one to run various load cases to optimize the utilization of the installed capacity around the data center.

3.1 Information Technology Infrastructure

IT infrastructure of the future will be challenged by the need to plan for growth and to meet unpredictable compute demand while making sure that the resource capacity is optimized. Companies such as HP and Sun, as a response, have introduced technologies that enable provisioning of compute resources based on need from a pool of resources [9]. A data center has been built in HP Laboratories in Palo Alto to study this approach. This data center is divided into three sections. The production section handles HP Laboratories' IT needs. The cluster in this section is comprised of 29 production racks -- six storage racks, two racks of network switch systems and 21 additional racks made up of 50 1U (1U = 1.75 inches) servers,

48 2U servers, 16 IA-64 Itanium 2U servers, 20 4U servers, 12 8U servers and 3 10U HP PA-RISC processor-based Unix servers. The second section is the HP Laboratories research racks area. It is made up of 17 racks housing 79 2U servers, 68 IA-64 Itanium 2U servers and a one 10U HP PA-RISC processor based Unix server. The third section is a high-density special projects area, made up of 15 racks housing two-way 520 1U IA-32 servers where 500 nodes are available for production.

3.2 Physical Design and Model of the Data Center

The HP Laboratories data center allows adaptability to customer needs. As an example, it has “smart” cooling features that enable the cooling resources to modulate thermodynamic and flow work. Thermodynamic work is the ability to extract heat and flow work is the change in mass flow of coolant supplied to the data center [1, 5]. The data center modeled in this study houses 66 server racks (configuration is subject to change and the number of racks is expected to increase in the near future) arranged in a hot and cold aisle configuration, as shown in figure 2.

The total data center area is 267 m² (2875 ft²) divided into three sections; South DC or production section, 126.7 m² (1364 ft²), houses 34 racks with a measured computational load of 136.5 kW. A Northeast DC or research section, 63.8 m² (687 ft²), houses 17 racks with a measured computational load of 89.3 kW. A Northwest DC or high-density section, 76.6 m² (824 ft²) houses 15 racks with a measured computational load of 144 kW. These numbers represent the operational load at a given time. For modeling purposes, the total heat load dissipated by the server racks of 369.8 kW is used and the total cooling capacity available from the CRAC units is up to 600 kW when all CRAC units are operational and running at

full capacity (100 kW per CRAC unit). The total flow rate provided by the CRAC units is 38.18 m³/sec (80900 CFM) based on 6.84 m³/sec (14500 CFM) from CRAC 1 thru CRAC 3 and 6.14 m³/sec (13000 CFM) from CRAC 4 and 5.76 m³/sec (12200 CFM) from CRAC 5 and CRAC 6. The CRAC flow rate and the thermodynamic extraction have to be captured in the model definition. The total energy consumption in data center cooling comprises the total work done. A chilled water-cooling coil in the CRAC unit extracts the heat from the air and cools it down to 20°C. The cooled air enters the data center through vented tiles located on the raised floor near the inlet of the racks.

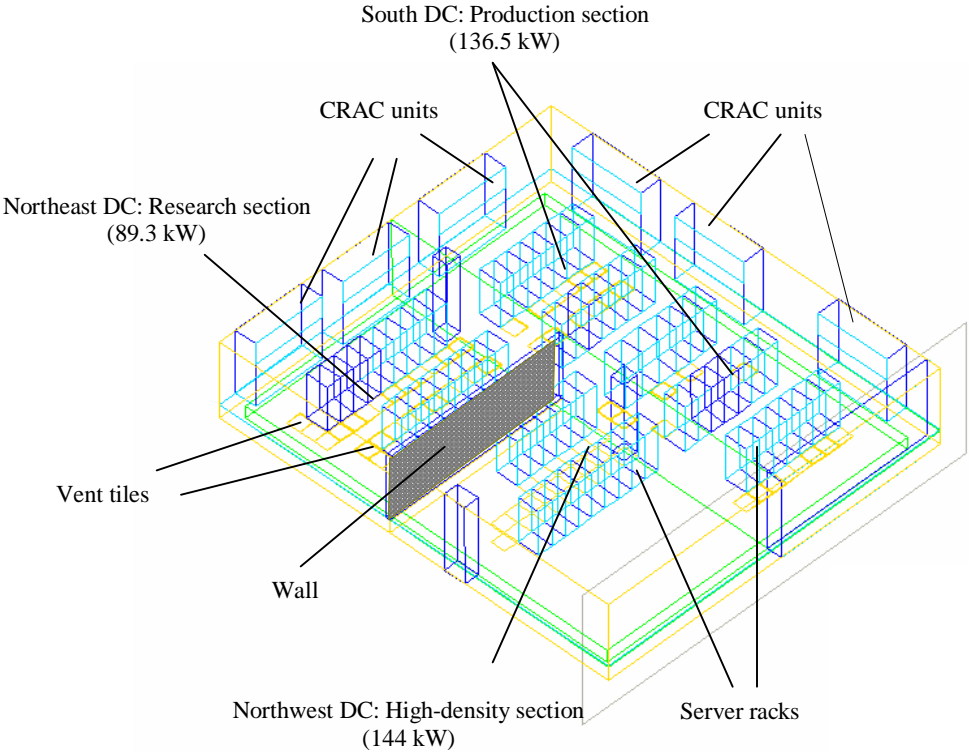


FIG. 2 Data Center Layout

In this paper, the principal features of the mathematical model are defined to be a three-dimensional model with negligible radiation effect. The motion of the fluid layers within the space of the data center is affected by fluctuating currents caused by mixing, entrainment and circulation in the computational domain and is represented using a turbulence model. Physically, turbulence increases mixing and the transfer of heat within the fluid and this is represented in the model as increased viscosity and thermal conductivity. The $k-e$ turbulence model that calculates the kinetic energy and its dissipation rate is used in this case.

3.4 Steady State Model

The model is built based on an actual Hewlett-Packard data center taking into account all of the data center details that affect air flow and temperature distribution including plenum size, chilled water pipes, cable trays, number of vent tiles and each vent tile's percentage opening, racks layout and each rack's computational load. The model defines servers as heat sources with their corresponding airflow rates. There are six CRAC units modeled with proper thermo-mechanical attributes supplying air at a given temperature. In this example, a 20 °C is set as the supply temperature. There are 74 vent tiles strategically distributed within the data center and are modeled appropriately with flow resistance term based on free area opening of 47% or 25%.

3.5 Transient Model

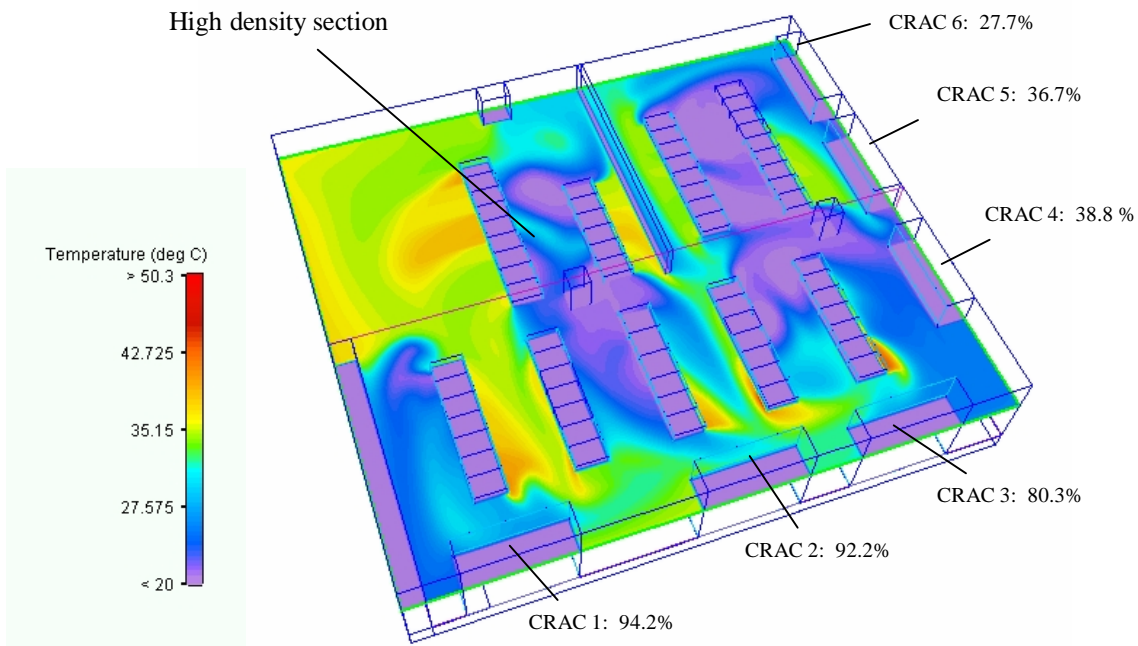
This model utilizes the steady state solution as the base solution and steps forward in time as the sudden change occurs. In transient simulations, the initial conditions are part of the problem specification and the model variables use the steady state solution as their initial conditions. The details of the model are the same as the steady state one. In the transient

calculations, the program solves the equations for a number of time steps with the time step is defined by the user. In this model, one CRAC unit is de-activated to simulate failure of the unit while maintaining the same work load in the data center. Monitor points were positioned in key areas within the model prior to solving. Values of all the variables at these defined points are graphed during the solution process which allows a dynamic assessment of the transient solution and a determination of the level of convergence. The total transient time specified is 300 seconds. Increasing time-steps are selected to capture the initial critical temperature changes just after the CRAC unit failure.

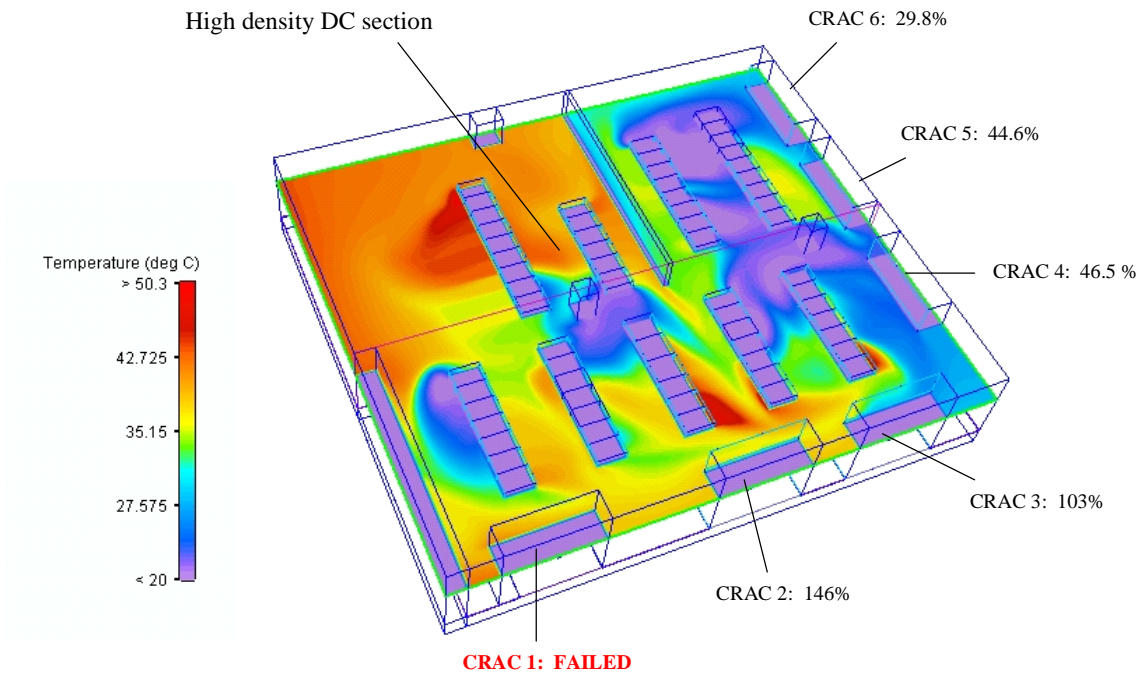
4. Results and Discussion

The steady state model is first created to simulate the thermal environment in the data center when all cooling resources are operating normally. This steady state model is then used as a reference state (initial conditions) to the transient model to simulate the environment when a sudden reduction in the data center cooling resources suddenly takes place. Fig. 3(a) shows the temperature distribution when all CRAC units are operating normally. The plot shows the capacity per CRAC unit where CRAC 1 and 2 are handling most of the load in the data center due to the load/racks layout relative to the CRAC units. In this case study, the contribution from CRAC units 5 and 6 to the high density area is minimal due to existence of a wall as shown in Fig. 2.

Fig 3(b) shows the temperature distribution in the data center when CRAC 1 has failed. The plot shows the high heat load accumulating the northwest high density section causing the temperature to rise. The plot also shows that failed CRAC unit was handling most amount of load compared to the other five CRAC units as previously shown in steady state model.



(a) Before CRAC unit failure



(b) After CRAC unit failure

FIG. 3. Temperature distribution before and after CRAC unit failure.

After failure of CRAC 1, previously separated warm air exhausted from the high-density racks in the Northwest section will have to travel across the room to reach CRAC units 2 through 4 warming up some of the cooled air along its path. The model also shows CRAC 1 failure leads to the mal-provisioning of CRAC 2 and CRAC 3 and to severe temperature rises in some regions (Northwest DC section) of the data center. These unacceptable environmental conditions could result in system failure in those regions.

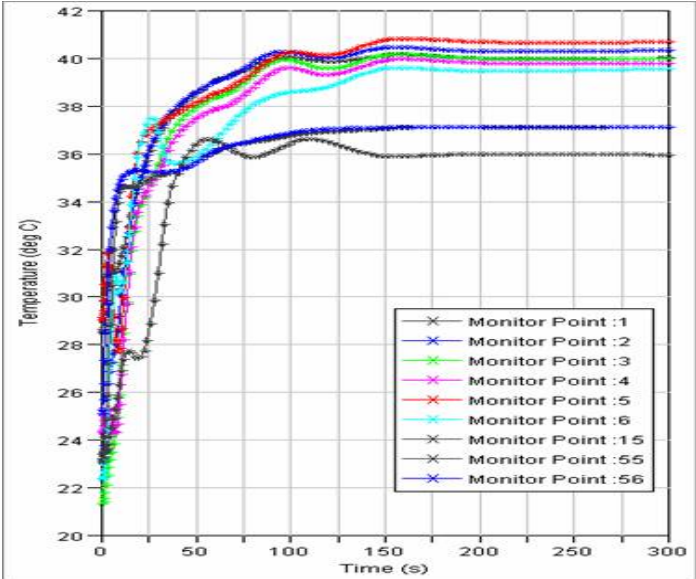


FIG. 4. Transient temperature solution profile.

Fig 4 show the transient temperature profile as the solution progressed to a satisfactory level of convergence. Temperature values in this plot are those of the monitoring points placed at the inlet of server racks. The placement of these monitoring points took into account the temperature distribution along the height of the rack inlets and defined the top of the rack as the best location. There are two ways to react to this failure: 1) changing the provisioning of

cooling resources (e.g. modifying air delivery location, temperature and flow rate) and/or 2) workload migration.

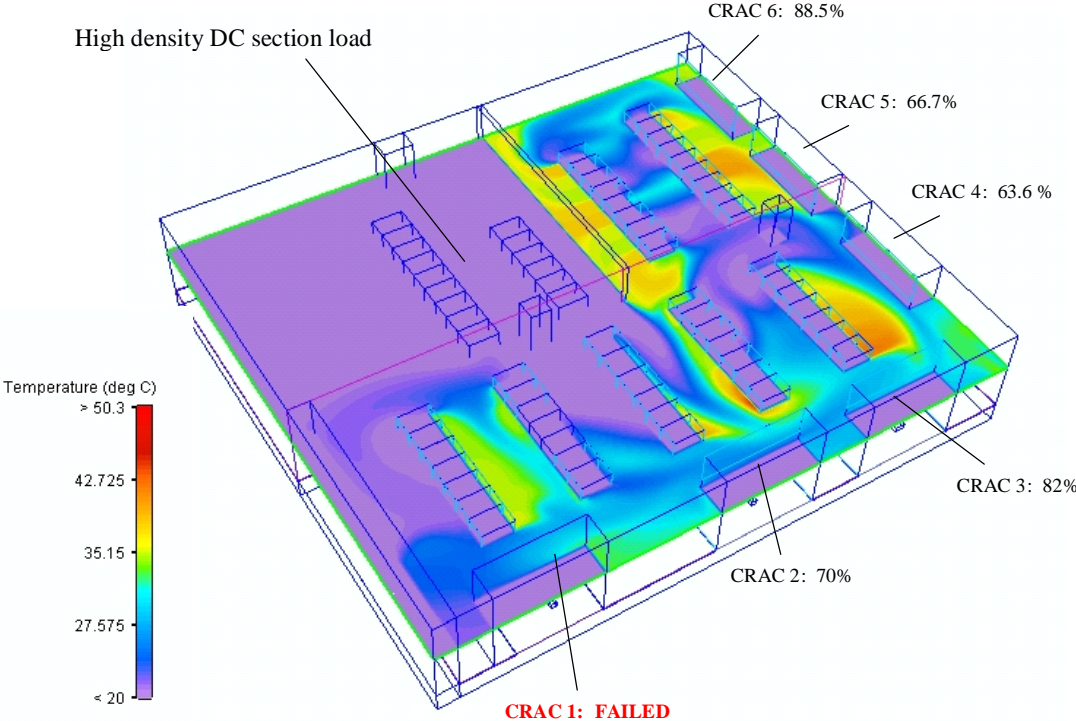
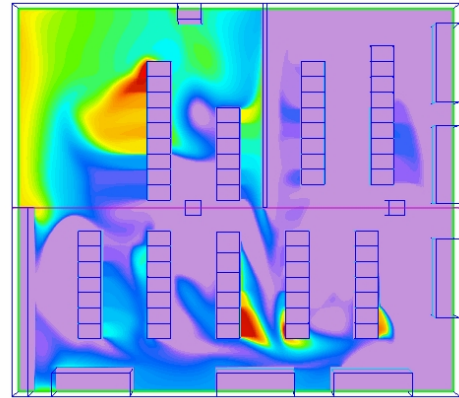
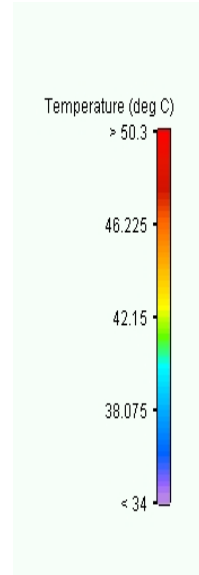


FIG. 5. Temperature Plane Plot after CRAC unit failure and load redistributed.

In this analysis, workload migration was simulated in the model by moving the entire load in the high-density area and redistributing it within the rest of the data center. This method assumes computational resources and cooling capacity is available at the time of the failure, as is the case in most data centers. Fig. 5 shows the data center can continue operation without application downtime with redistributed computational loads and a different provisioning of CRAC units. Fig 6 shows a top plan plot of the temperature distribution for both cases showing the sections where temperature exceeded 34°C.

6a. Temperature distribution above 34 C after CRAC unit failure and before computational load redistribution.



6b. Temperature distribution above 34 C after CRAC unit failure and computational load redistribution.

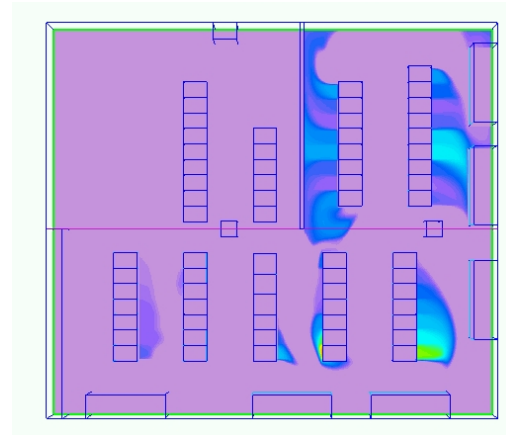


FIG. 6. Temperature Plane Plot after CRAC unit failure before and after load redistribution.

Summary and Conclusions

A computational fluid dynamics model of a high-density data center is created to provision the cooling resources and determine the effect of failure in cooling infrastructure on the systems. Steady state and transient models are created and the effect of the CRAC unit failure on the data center servers is analyzed. The transient model shows that the rack inlet temperature will increase to an unacceptable level in less than 80 seconds after a CRAC unit failure for a given distribution of high power density racks. Mitigation of this failure can be planned by proper analysis, design and “smart” application of redundant cooling systems.

Besides “smart” application of redundancy at air conditioning level, active management of power and migration of workload can be utilized to avert a disaster.

The salient points of the paper can be summarized as follows:

- § Today’s servers enable high compute density in data center and current design approaches must be augmented with detailed thermo-fluids analyses. Providing sufficient cooling resources based on the total compute (heat) load in the data center is not enough, as local high power density can lead to mal provisioning of CRAC units.
- § Modeling the thermo-fluids behavior in the data center can prove useful to determine uptime, create policies and procedures to mitigate failure and enable reliable operation.
- § A viable approach to mitigate failure of a cooling resource is to migrate the compute loads within the data center and maintain specified inlet temperature to enable reliable operation.
- § Thermo-Fluids policies that enable workload migration and lead to efficient and reliable use of cooling resources are being developed [6, 7].

Nomenclature

div	mathematical operator
E	energy, Q
e	specific internal energy, Q/M
g	gravitational acceleration, M/t ²
grad	mathematical operator
i, j	symbols denote any of the space coordinate subscripts.
k	thermal conductivity, Q/tLT
P	system pressure, F/M ²
S	source term
u	velocity in the x-direction, M/t
v	velocity in the y-direction, M/t
w	velocity in the z-direction, M/t
x	space coordinate in x-direction, L
y	space coordinate in y-direction, L
z	space coordinate in z-direction, L

Greek Symbols

•	kinematic viscosity, M ² /t
t	viscous shear tensor, F/L ²
r	density, M/L ³
∇	mathematical operator
m	dynamic viscosity, Ft/L ²
S	stress tensor, F/L ²
l	thermal conductivity, Q/tLT

Dimension Symbols

F	force
L	length
M	mass
Q	energy
t	time
T	temperature

References

- [1] Patel, C.D., Sharma, R.K, Bash, C.E., Beitelmal, A, “Thermal Considerations in Cooling Large Scale High Compute Density Data Centers”, ITherm 2002 - Eighth Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, San Diego, CA, May 2002
- [2] Patel, C.D., Bash, C.E., Belady, C., Stahl, L., Sullivan, D., July 2001, Computational fluid dynamics modeling of high compute density data centers to assure system inlet air specifications, Proceedings of IPACK’01 – The PacificRim/ASME International Electronics Packaging Technical Conference and Exhibition, Kauai, Hawaii
- [3] Sharma R, Bash, C.E, Patel, C.D, June 2002, “Dimensionless Parameters for Evaluation of Thermal Design and Performance of Large Scale Data Centers”, AIAA-2002-3091, American Institute of Aeronautics and Astronautics Conference, St. Louis.
- [4] Bash, C.E., Patel, C.D., Sharma, R.K., “Efficient Thermal Management of Data Centers – Immediate and Long-Term Research Needs”, International Journal of Heat, Ventilating, Air-Conditioning and Refrigeration Research, April 2003
- [5] Patel, C.D., Bash, C.E., Sharma, R.K, Beitelmal, A and Friedrich, R.J., “Smart Cooling of Datacenters”, Proceedings of IPACK’03 – The PacificRim/ASME International Electronics Packaging Technical Conference and Exhibition, Kauai, HI, IPACK-35059, July 2003
- [6] Sharma, R., Bash, C.E, Patel, C.D, Friedrich, R.S, Chase, J, “Balance of Power: Dynamic Thermal Management for Internet Data Centers”, Hewlett-Packard Laboratories Technical Report: HPL-2003-5.
- [7] Patel, C.D., Sharma, R.K, Bash, C.E. and Graupner, S., 2003 “Energy Aware Grid: Global Workload Placement based on Energy Efficiency”, IMECE 2003-41443, 2003 International Mechanical Engineering Congress and Exposition, Washington, DC.
- [8] Flomerics Incorporated, <http://www.flovent.com>, Flovent version 4.2.
- [9] Graupner, S., Chevrot, J.-M., Cook, N., Kavanappillil, R., Nitzsche, T., “Adaptive Control for Server Groups in Enterprise Data Centers,” fourth IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2004), Chicago, Illinois, April 19-22, 2004.