# Thermal Efficiency: Facebook's Datacenter Server Design

By John Parry, Industry Manager, Mentor Graphics

**Large-scale datacenters consume megawatts in power and cost hundreds of millions of dollars to equip. Reducing the energy and cost footprint of servers can therefore have substantial impact.**

Web, Grid, and Cloud Servers in particular can be hard to optimize, since they are expected to operate under a wide range of workloads. For its first datacenter in Prineville, Oregon, Facebook set out to significantly improve its power efficiency, cost, reliability, serviceability, and environmental footprint. To this end, many dimensions of the datacenter and servers were redesigned, using a holistic approach. This article is abstracted from the Facebook paper "High-efficiency Server Design," which was presented at the 2011 ACM Conference on Supercomputing, and focuses on this server design, combining aspects of power, motherboard, thermal, and mechanical design. In this article we have looked at the thermal aspects in isolation. In the full paper, Facebook calculated and confirmed experimentally that its custom-designed servers can reduce power consumption across the entire load spectrum while at the same time lower acquisition and maintenance costs. The design does not reduce the servers' performance or portability, which would otherwise limit its applicability. Importantly, the server design has been made available to the open source community via the Open Compute Project, a rapidly growing community of engineers around the world whose mission is to design and enable

This article was originally published in Engineering Edge Vol. 3 Iss. 1


Mentor Graphics
Mechanical Analysis

the delivery of the most efficient server, storage and datacenter hardware designs for scalable computing.In the past decade, we have witnessed a fundamental change in personal computing. Many of the modern computer uses such as networking and communicating; searching; creating and consuming media; shopping; and gaming—increasingly rely on remote servers for their execution.

The computation and storage burdens of these applications has largely shifted from personal computers to the datacenters of service providers such as Amazon, Facebook, Google, and Microsoft. These providers can thus offer higher-quality and larger-scale services, such as the ability to search virtually the entire internet in a fraction of a second. It also lets providers benefit from the economies of scale and increase the efficiency of their services.

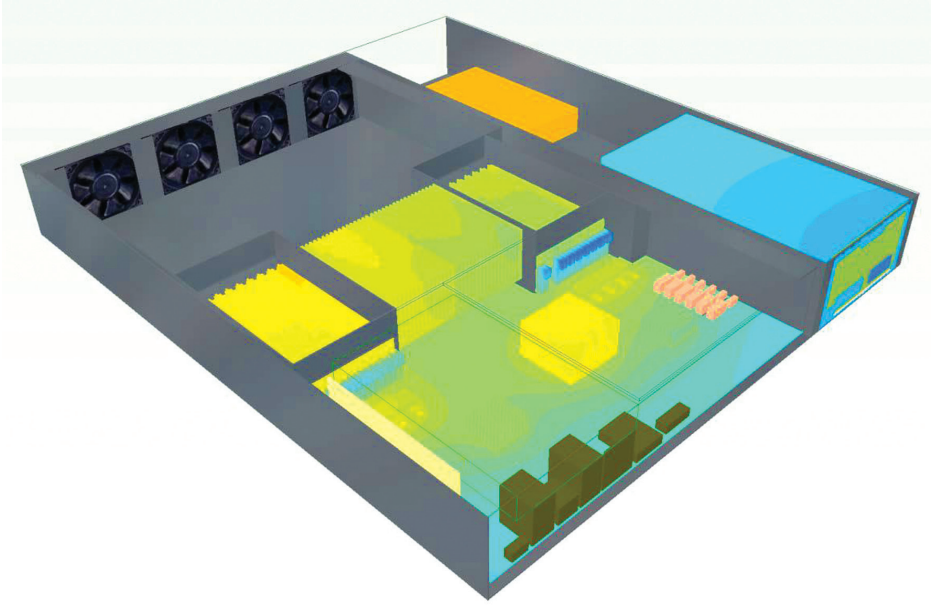

**Figure 1.** FloTHERM isometric view of thermal design shows chassis, motherboard (with dual processors and memory slots side-by-side), fans, and the hard-disk drive (HDD) behind the PSU. The temperature range here assumes an inlet temperature of 27°C. The air duct on top is elided for visualization purposes.

As one of these service providers, Facebook leased datacenters and filled them with commodity servers. This choice made sense at small to medium scale, while the relative energy cost is still small and the relative cost of customization outweighs the potential benefits. As the Facebook site grew to become one of the world's largest, with a corresponding growth in computational requirements, they started exploring alternative, more efficient designs for both servers and datacenters.

## Thermal Design

The goal of server thermal design is to cool down the hot components to their operating temperatures with a minimal expenditure of energy and component cost. The typical mechanism used to cool servers at the datacenter level is to cool air at large scale and push it through the servers using their internal fans. The cool air picks up heat from the server components, exits from the server outlet, and is then pushed back to the atmosphere or chilled and recirculated.

More efficient cooling is achieved with air containment in aisles, with the front (or inlet), side of the server facing the "cold aisle" and the back facing the "hot aisle." Yet another technique to improve cooling efficiency is to create an air-pressure differential between the aisles using large datacenter fans. In this case the specific design goal was to be able to cool the upcoming datacenter without chilling the outside air almost year round by allowing effective server cooling even with relatively high inlet air temperature and humidity. To achieve this goal, a more effective design was needed for heat transfer than currently used in the commodity servers.

Improving airflow through the server is a key element here: when internal server components impede airflow, more cooling energy is expended (for example, by faster fans, cooler inlet air, or higher air pressure). One technique by which improved airflow is achieved in the chassis is to widen the motherboard and spread the hot components side by side, not behind each other. The hottest components—processors and memory—were moved to receive the coldest air first, by locating them closer to the air inlet than in the typical back-mounted motherboard.

Another modified dimension was the server height: given a relatively constant rack height (for servicing purposes), a taller server reduces cooling energy but also the rack's computational density. Calculations found that the optimal server height to maximize the compute-capacity per cooling-energy ratio to be the uncommon 1.5U height with large-surface-area heat sinks. This height also allows for an air duct that sits on top of the motherboard and "surgically" directs airflow to the thermal components in parallel heat tracks, reducing leaks and air recirculation inside the chassis. Obstructions to airflow are kept to a minimum, decreasing the number of fans required to push the air out (Figure 1).

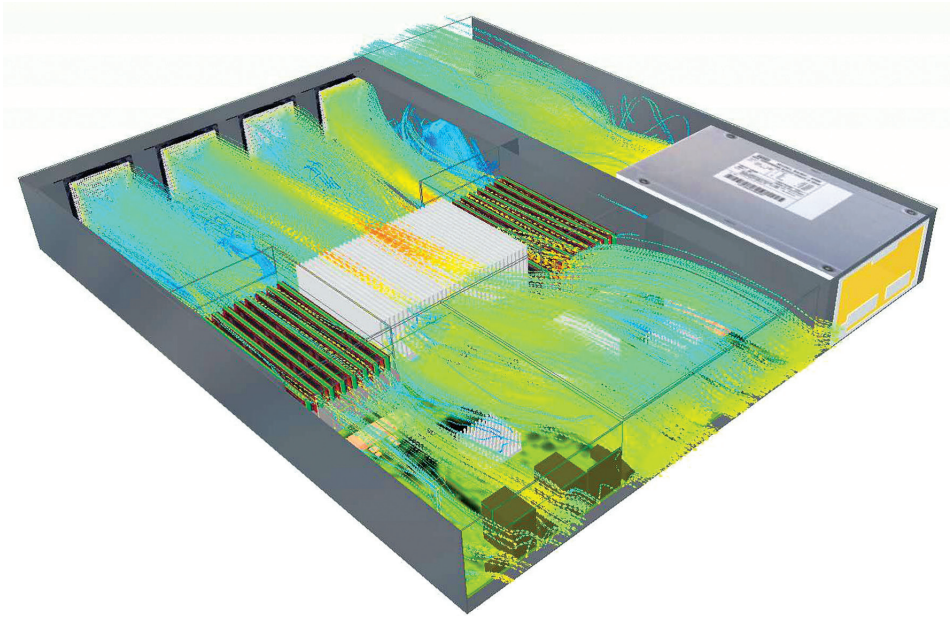

**Figure 2.** FloTHERM CFD simulation of airflow speed at minimum continuous fan speed



This article was originally published in Engineering Edge Vol. 3 Iss. 1

And since the high-efficiency PSU generates less than 20W of waste heat under load, the HDD remains well within specified temperature operating range even behind the PSU. Contrast this with typical server designs that locate the HDD in the front of the chassis to meet its cooling requirements. Also reduced is the amount of airflow required through the system to keep it cool—up to half the volume flowrate compared to standard 1U servers, for the same inlet to-outlet temperature difference (Figure 2).

This low requirement, combined with smart fan-speed controllers, results in fans that spin at their minimum continuous speed nearly year-round, depending on ambient temperature and workload.

An additional advantage of this low speed, continuous operation is a longer expected fan lifetime compared to the typical fan's start-stop cycles, leading to overall improved server reliability. It also naturally translates to lower power and operating costs for server cooling—approximately 1% of the total server power—compared to the more typical 10% in commodity servers. Somewhat surprisingly, even the CAPEX of the server's cooling components alone is about 40~60% lower than a typical server, depending on OEM component pricing. The two main reasons for this improvement are the use of thinner fans (owing to the reduced airflow) and simpler heatsinks without a heat pipe (owing to the larger surface area). Closing the cycle, these efficiency gains carry forward to the datacenter level as well. The server is capable of working reliably at air inlet temperatures of 35°C and a relative humidity of 90%, exceeding the most liberal ASHRAE recommendations for datacenter equipment. In practice, this allows Facebook's datacenter to be cooled almost exclusively on free (outside) air, relying on infrequent evaporative cooling instead of chillers only on particularly hot days.

## Methodology

Facebook have evaluated the power, thermal and performance properties of a prototype of the new design against two commodity servers. Both commodity servers are a common off-the-shelf product from two major OEMs, with dual Xeon X5650 processors, 12GB DDR3 ECC memory, on-board Gigabit Ethernet, and a single 250G SATA HDD in a 1U standard configuration. The first server, "Commodity A," is widely deployed in the leased datacenters for Facebook's main Web application. The second server, "Commodity B," is a three-year-old model that was updated to accept the latest generation processors. To ensure a fair comparison, the exact same CPUs, DIMMs, and HDD unit are used in turn, moving them from server to server. The only differing components between the three servers were therefore the chassis, motherboard, fans, power supply, and power source (208V ac/277V ac).

## Thermal Efficiency

Thermal efficiency is another important element of the total cost of ownership (TCO), both in terms of cooling energy in the server (fan energy) and in the datacenter. The thermal design is based on a spread and unpopulated board placed in a 1.5U pitch open chassis, and employs four high-efficiency custom 60 × 25mm axial fans. In contrast, the commodity servers use a thermally shadowed, densely populated 1U chassis with six off-the-shelf 40×25mm fans. To evaluate the thermal efficiency, each server was placed in a specially-built airflow chamber that can isolate and measure the airflow through the server, expressed in cubic-feet per-minute (CFM). The measured CFM value was also confirmed analytically by measuring the server's AC power and air temperature difference between inlet and outlet. The servers are loaded with an artificial load resembling Facebook's production power load (around 200W, with leakage power at less than 10W), while maintaining the constraint that all components remain within their operating thermal specifications. The results for the prototype (Figure 3) show a significant improvement. For a typical 7.5MW datacenter, this reduced airflow translates to a reduction of approximately 8~12% of the cooling OPEX. More importantly, it enables free air cooling to be used for the datacenter.

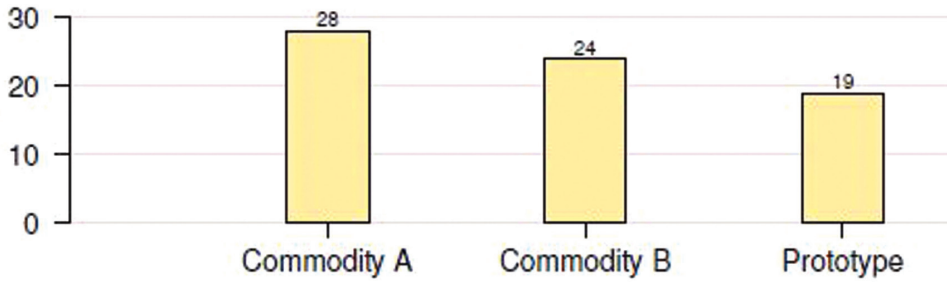


**Figure 3.** Airflow comparison (in CFM) at 200W

## Conclusions

This new server design measurably reduces TCO without reducing performance. The customized server design can:

1. Reduce operating and cooling power (e.g. efficient power conversions, higher-quality power characteristics, fewer components, thinner and slower fans, improved airflow).
2. Lower the acquisition cost and server weight (e.g. fewer and simpler components, lower density, fewer expansion options).
3. Cut costs on supporting infrastructure (e.g. no centralized UPS, no PDUs, no chillers).
4. Increase overall reliability (e.g. fewer and simpler components, distributed and redundant batteries, smooth normal / backup transitions, staggered HDD startup, slower fans).
5. Improve serviceability (e.g. all-front service access, simpler cable management, no extraneous plastics or covers).

At large scale, this design translates to substantial savings. Facebook calculate that over a three year period, these servers alone will deliver at least 19% more throughput, cost approximately 10% less, and use several tons less raw materials to build than a comparable datacenter of the same power budget, populated with commodity servers. When matched with a corresponding datacenter design (including all aspects of cooling, power distribution, backup power, and rack design), the power savings grow to 38% and the cost savings to 24%, with a corresponding power usage effectiveness (PUE) of ≈ 1.07.

## Reference:

[1]. Eitan Frachtenberg, Ali Heydari, Harry Li, Amir Michael, Jacob Na, Avery Nisbet, Pierluigi Sarti, Facebook. High-Efficiency Server Design, 2011

This article was originally published in Engineering Edge Vol. 3 Iss. 1


Mentor Graphics

Mechanical Analysis

This article originally appeared in Engineering Edge Vol. 3 Iss. 1

Download the latest issue:
www.mentor.com/mechanical/products/engineering-edge



# ENGINEERING EDGE

Accelerate Innovation with CFD & Thermal Characterization

## You might also be interested in...

**Facebook Datacenter Design: Thermal Efficiency in Datacenter Server Design**

**Stanley Black & Decker Power Tool Designs**

**Voxdale and Flanders' DRIVE Deliver Automotive Innovation**

**Seiko Epson: Empowering Engineers since 1989**

**Electronic Cooling Solutions Inc.: Comparing Tablet Natural Convection Cooling Efficiency**

**Cofely Fabricom E&E: A Simulation Driven Approach to a Biomass Furnace Upgrade**

**Download the latest issue:**
**www.mentor.com/mechanical/products/engineering-edge**