Holistic Datacenter Design in the Open Compute Project

Eitan Frachtenberg
Facebook

Facebook's Open Compute Project lets the community benefit from and contribute to improvements in power and water usage effectiveness, cost, and operation.

Datacenters have fundamentally changed the way we compute. Most things we do with a computer nowadays—networking and communicating, analyzing data, searching the Web, creating and consuming media, shopping and gaming—increasingly rely on remote servers for execution.

These applications' computation and storage burdens have largely shifted from client computers to datacenters of service providers, including Amazon, Facebook, Google, and Microsoft. These providers can offer users higher-quality and larger-scale services, such as searching virtually the entire Internet in a fraction of a second.

Concentrating data and computation in datacenters is beneficial not only to users but also to service providers. By consolidating infrastructure, codesigning hardware and software, and taking advantage of economies of scale, providers can customize and optimize an entire datacenter for only a small number of applications as if it were a single warehouse-scale computer, thereby reducing the total cost of ownership and improving product quality.

The general public hasn't known much about these optimizations. Most companies have regarded the optimizations as a competitive advantage and refused to share how they squeeze higher efficiencies out of their datacenters or to even acknowledge their locations. This changed in 2011, when Facebook launched the Open Compute Project (OCP), which releases the specifications of its custom-built datacenter and server designs (www.opencompute.org).

OCP's main goal is to foster collaboration and technology improvements in datacenter design. This collaboration could consequently reduce not only Facebook's cost and energy footprint but the entire industry's as well. By now, dozens of companies have joined as official members of the project, some already designing or opening their own components to work with OCP. What makes the nascent OCP so compelling to these industry stalwarts? In a word, efficiency.

HOLISTIC DESIGN DRIVES EFFICIENCY

The key to the OCP design (and other less transparent datacenters) is that codesigning all elements of the datacenter, instead of acquiring them separately on the commodity market or designing them piecemeal, can yield significantly higher efficiencies while maintaining equivalent performance.

OCP's specifications run the gamut from the voltage regulator on a motherboard to the datacenter building's shape and climate requirements. Some design decisions constrain other choices, whereas others enable new design freedom. Starting from a nearly blank slate and redesigning a datacenter holistically, without first worrying about compatibility with previous choices and standards, can usher in radical new designs and, consequently, efficiencies.

Perhaps the best example of this is the way power is distributed through the datacenter. We chose the nonstandard voltage of 277 Vac (volts of alternating current), which necessitated redesigning the server power supplies to accept the higher

0018-9162/12/$31.00 © 2012 IEEE Published by the IEEE Computer Society

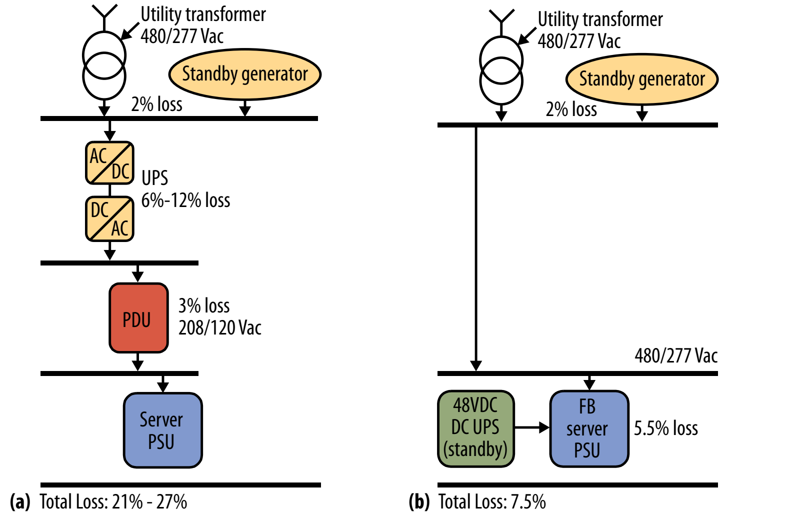

Figure 1. Datacenter power distribution: (a) a typical datacenter and (b) in the Open Compute Project design.

voltage, as well as an entire new distributed backup scheme. On the other hand, this strategy helped to eliminate most of the wasteful voltage conversions of the traditional datacenter; and buying, operating, and maintaining the new backup scheme is more cost-efficient.

As Figure 1 shows, in a typical datacenter, power comes into the building at 480 Vac from the utility transformer (277 Vac lines-to-neutral), and is converted to direct current (DC) to charge an online uninterruptible power supply (UPS). It's then converted back to 480 Vac and is distributed through power distribution units (PDUs), where it's converted again to a standard 120 Vac, all the way to the server power supply units (PSUs).

These conversions create a power loss of about 21 to 27 percent, depending on implementation. In the OCP design, power is distributed at 277 Vac all the way to the server PSUs and to offline battery backup with no conversions, resulting in a total power loss of 7.5 percent or less. It does, however, require a custom-designed PSU and backup scheme.

Thermal design

Another example of how we literally redesigned the datacenter from the ground up is in its thermal design. Whereas traditional datacenter buildings start with a raised floor for the cooling infrastructure to push chilled air up to the data hall. In our Prineville, Oregon, datacenter, we decided to forgo most of the expensive, power-hungry cooling infrastructure altogether, including ductwork, chillers, cooling towers, and associated pumps.

Cooling the IT equipment is a crucially important function (and often a major cost) in the operation of any datacenter, and ours is no exception; but we saw an opportunity for improvement based on two basic insights.

First, traditional datacenters chill and recycle the same air over and over, when in reality outside air is almost universally cooler than the inside exhaust air; so, meeting operational specifications is less costly.

Second, even these operational specifications are typically overly conservative, when in reality most IT equipment can operate at higher temperature and humidity ranges than even the most liberal specification from the American Society of Heating, Refrigeration, and Air-Conditioning Engineers (ASHRAE; www.ashrae.org).

Building design

The datacenter is a two-storied building, with the first floor holding the data hall and office space, along with the receiving yard and storage area. (E. Frachtenberg et al., "Thermal Design in the Open Compute Datacenter," to be published in *Proc. 13th IEEE Intersociety Conf. Thermal and Thermomechanical Phenomena in Electronic Systems* [ITHERM 12], IEEE, 2012).

The data hall was codesigned with the IT equipment and air pathways to minimize airflow restriction, reducing the amount of energy required to move air around. There's a large plenum above the data hall for hot return air. The second floor is a built-up mechanical penthouse that holds an evaporative cooling and humidification system.

As Figure 2 shows, outside air enters on the upper floor through the left side of the building, and pollutants are filtered out in the filter room. In the evaporative cooling and humidification room, the air is misted with a fine spray of purified water to control the temperature and humidity as needed.

The air then proceeds through a fan wall, which is usually the only moving mechanical element in the cooling system. The cold air naturally descends from the supply corridor to the data hall, where it's pressured through the IT equipment, picking up heat, and out the back to the contained hot aisles.

The air then ascends to the exhaust corridor. From there, some air is used to heat up the office space or incoming air as needed, while the rest is vented back to the atmosphere on the top-right side of the building.

The servers are also designed to work together with the building's cooling system for maximum effectiveness. Some of the design innovations include

- nonstandard rack and motherboard dimensions to allow for

larger, more efficient heat sinks and server fans;
- side-by-side arrangement of hot components to allow near-laminar airflow; and
- high-efficiency electrical components that release less heat.

To improve airflow and reliability and to reduce cost, power consumption, and resource waste, we also removed many extraneous plastic covers, bezels, and functional components from the servers that seemed irrelevant to our workload, (E. Frachtenberg et al., "High Efficiency Server Design," *Proc. IEEE/ACM Conf. Supercomputing* [SC 11], ACM, 2011, article 27).

BEYOND POWER USAGE EFFECTIVENESS

Although these design changes require significant engineering investment and possibly a departure from the commodity market, their efficiency benefits add up quickly. For example, in the first OCP datacenter in Prineville, the power usage effectiveness (PUE) ratio, defined as the total power entering the building divided by the total power entering the IT equipment, averaged 1.08 over the first year of operation—the lowest reported number that takes into account all of the datacenter's grid power.

At this point, with only 8 percent wasted energy, further improvements to the PUE metric are no longer dramatic. Moreover, the PUE metric has been widely criticized, for example, because it favors less efficient designs where more energy is spent on the server fans, as opposed to the datacenter's fans. In fact, OCP servers consume 16 to 40 percent less energy than the commodity servers they replaced at equivalent performance, penalizing the PUE metric.

The time is ripe to look at datacenter efficiency more holistically rather than focusing on PUE. Among additional metrics, we have also measured the total acquisition cost (capex) and operation cost (opex) of our OCP datacenter and found them to be, respectively, 24 percent and 38 percent lower than for leased datacenters.

Figure 2. Rendering of airflow through the OCP datacenter.

Our calculated water usage effectiveness (WUE) is 0.31 liter/kilowatt hour, which we believe compares favorably with other state-of-the-art datacenters—although very little has been published about WUE. Additionally, we found that OCP servers use far less material to build—about six pounds' worth—and are easier and faster to maintain than commodity servers.

All these efficiencies are intertwined. For example, higher-efficiency servers emit less heat and require less power and water for cooling, which contributes to lower PUE and WUE. Cost is also tightly linked to efficiency. Obviously, the opex is directly affected by the amount of overall power used in the datacenter, so lower PUE translates to a lower opex. But efficiency also affects the capex, sometimes indirectly. A lower PUE means less overall power, which requires less power distribution equipment to provision and lower backup power requirements, reducing overall cost.

In some cases, high-efficiency components and a distributed backup power solution can actually be cheaper and simpler. In addition, fewer components also increase reliability, further reducing cost and increasing efficiency.

Previous articles have already pointed out some of the principles outlined here—including in this column. For example, in "Saving the World, One Server at a Time" (*Computer*, May 2011, pp. 91-93), Parthasarathy Ranganathan and Jichuan Chang discussed holistic datacenter design, understanding energy holistically, and dematerialization in the datacenter. But prior to OCP, these principles haven't been gathered in one successful implementation, and then shared freely with the world to iterate upon.

Our hope is not only that other datacenter designers will adopt OCP for their own needs, which already could reduce the environmental footprint of their datacenters, but that they also will continuously improve upon the design and contribute to it, amplifying the beneficial effect many times over.

***Eitan Frachtenberg** is a research scientist at Facebook. Contact him at etc@fb.com.*

Editor: Kirk W. Cameron, Dept. of Computer Science, Virginia Tech; greenit@computer.org