

Data Center

Backblaze Completes 500 Petabyte Data Center

BY JOHN RATH ON FEBRUARY 5, 2014



Rows of storage units inside the new Backblaze data center in the Sacramento market. (Photo: Backblaze)

Online backup provider **Backblaze** has completed a 500 petabyte data center, at the Sungard Availability facility just outside of Sacramento, California, the company said in a [blog post](#). After outgrowing its 40 Petabytes of storage in a caged facility in Oakland, the company set out in 2012 to find a new home. After reviewing proposals from all over the nation, Sungard was selected, and the staff went to work installing the company's signature Storage Pods. The data center also has SAS 70 Type II and ISO 9001 certifications and is PCI-DSS compliant.

The Sacramento data center has been quietly receiving customer data, and by September of last year all new customer accounts were being serviced there. Backblaze expects to store 500 petabytes of customer data at the new facility.

The Sacramento data center has been quietly receiving customer data, and by September of last year all new customer accounts were being serviced there. Backblaze expects to store 500 petabytes of customer data at the new facility.

The extremely cost-efficient Backblaze Storage Pod 1.0 caught the interest of many in 2009, and its current [3.0 Pod](#) packs 180 terabytes in a re-designed 4U chassis with



many upgraded components. Having shared their design ideas on the Storage Pod architecture, other companies such as [Netflix](#) were inspired to design their own custom storage appliances. The complete story of Storage Pod 3.0, the architecture, specs, and economics are in [this](#) February 2013 blog post.



For most tech companies, the process of selecting a data center partner and building out the facility itself is shrouded in secrecy. All parties to the transaction are usually bound by strict non-disclosure agreements.

Not so for [Backblaze](#). The San Mateo, CA., company, which offers a [cloud-based data backup service](#), is in the market for a new data center and isn't shy about discussing it. In fact, it's posting a link to its RFP on [its blog](#).

"Many companies have this knee-jerk reaction about disclosure. They worry that things they say will somehow be used against them," Backblaze CEO Gleb Budman told me in a recent interview. Backblaze has been open about some other things, too. It shared information about [how it builds the storage pods](#) that run the company's [cloud-based data backup service](#). It also talked about why it didn't initially seek venture funding and also about [why it changed its mind](#).

"There may be some things that should be proprietary for a very specific reason, but if we don't see a reason not to share, we share," Budman said



The RFP lays out exactly what the company is looking for in terms of initial physical, power, and network requirements and information on what it thinks it will need for expansion over the course of the next 3 to 5 years. For example, it expects to add 1 to 1.5 cabinets per month over a 36 month period. That's because Backblaze is nothing if not growing. The company manages

40 petabytes of data now and adds about 2 petabytes per month.

Interestingly, while Apple(s aapl), Google(s goog), Amazon(s amzn) remain incredibly tight-lipped about their data center build outs — our own Katie Fehrenbacher can attest to that after her [data center road trip](#) — there's also a countervailing trend pushing for more openness about data center gear. The biggest example is the Facebook(s fb) initiated [Open Compute Project](#), which pushes for standards-based, energy-efficient data center hardware. These companies may not want you in their data center, but they're now willing to share some best practices about the hardware you should run in yours.

Backblaze's 8-page RFP asks prospective data center providers what they offer in terms of physical security, and SAS 70 compliance and other regulatory certifications.

It's an interesting document that could act as a template for other companies that are now in the market for a new data center.

DIY data centers: Backblaze reveals secrets of cloud storage

REGINA HOPE SINSKY JULY 20, 2011 11:02 AM

If Hollywood ever needs a cloud storage facility for a movie scene, they need to call the folks at online storage and backup

company [Backblaze](#). You can just picture Tom Cruise, playing a secret agent dressed in black, sneaking around the rows of shiny



red pods in the white storage facility (half a city block long). This is what cloud storage should look like, and the company is now sharing their secrets.

You too can make a 135-terabyte, 4U server for \$7,384.

Today Backblaze posted a new blog full of information most companies keep hidden about building and running a 15+ petabyte cloud storage farm.

"When we first open-sourced the storage pod design in 2009, it was the first time it had ever been done," says Gleb Budman, co-founder and CEO of Backblaze, in an interview with VentureBeat. "500,000 people read the blog post and hundreds of companies around the world have since built the storage pods for their own purposes."

Backblaze calls those 135-terabyte, 4U servers “storage pods.” They are self-contained units, composed of metal cases with commodity hardware inside, all designed to put storage online. The company shares images of a half-assembled pod. The cost of the hard drives dominates the price of the overall pod and the system is made entirely of commodity parts. For more background, read the original blog post.

Today’s Backblaze post includes details on how to make a version 2.0 storage pod, data on the total cost of ownership, the impact of heat on drives and more. This is one of the longest blog posts you will read today. Even if you aren’t interested in building a cloud farm, Budman’s writing is extremely novice-friendly. It provides some interesting insight to a technology most of us use — unconsciously — but few will ever see or think about.

“You’re welcome to use the design” is a phrase rarely uttered these days in Silicon Valley. Budman has it right there in his blog, in writing.

“We benefitted greatly from many other open-source projects and the efforts that thousands of people put into them,” he explains to VentureBeat. “We wanted to share some of that good karma back.”

This openness is unusual for a hardware company. Imagine Apple inviting people to fully examine and contribute to its products.

“We see ourselves as a software/service company at heart, trying to build the easiest online backup service available,” Budman says. “We built the hardware out of necessity because we could not find anyone selling the hardware we needed. Most of the storage servers available at the time cost more than \$1,000 per terabyte at a time when drives were less than \$100 per terabyte. We see our

secret sauce as the backup technology on the client and software technology for the cloud."

Budman says he gets feedback from "helpful kindred pod builders," including Justin Stottlemeyer from digital photo storage and sharing company Shutterfly and another unidentified person who setup the Google Group "OpenStoragePod." Budman says hundreds of individuals have emailed ideas and feedback to Backblaze over the last year and a half. Having this much conversation about a piece of hardware is unique in the competitive cloud industry.

With data center space (and power) at a premium, minimizing storage cost means maximizing density. Backblaze's new pods store twice as much data in the same space as the old ones. One datacenter rack containing 10 pods costs Backblaze about \$2,100 per month to operate (the price is divided into thirds for physical space rental, bandwidth and electricity). Doubling the density saves the company half of the money spent on both physical space and electricity. The new pod cabinets squeeze one petabyte into three-quarters of a single cabinet for \$56,696.

So how does Backblaze make money? The company's business is to provide an unlimited online backup service for about \$4 a month (if you buy for two years.) There aren't any hidden caps on storage and Backblaze automatically backup all data, including on external drives.

Backblaze is going against the Silicon Valley grain in another way: The company is employee owned. It has not taken any venture funding.

"If we raised VC funding, we could theoretically burn money to start and hope to

figure it out later,” says Budman. “Since we decided to bootstrap the company, we needed to ensure the business became cash-flow positive quickly. That forced us to innovate on the cost of storage.”

The company is now profitable and cash-flow positive, Budman says.

Budman says that to stay profitable, they have to keep costs and expenses low. He says that one of the hidden datacenter costs is the salary of employees who deploy pods, maintain them, replace bad drives with good, and generally manage the facility.

“Backblaze has 16 petabytes and growing, and we employ one guy whose fulltime job is to maintain our fleet of 201 pods, which hold 9,045 drives,” says Budman.

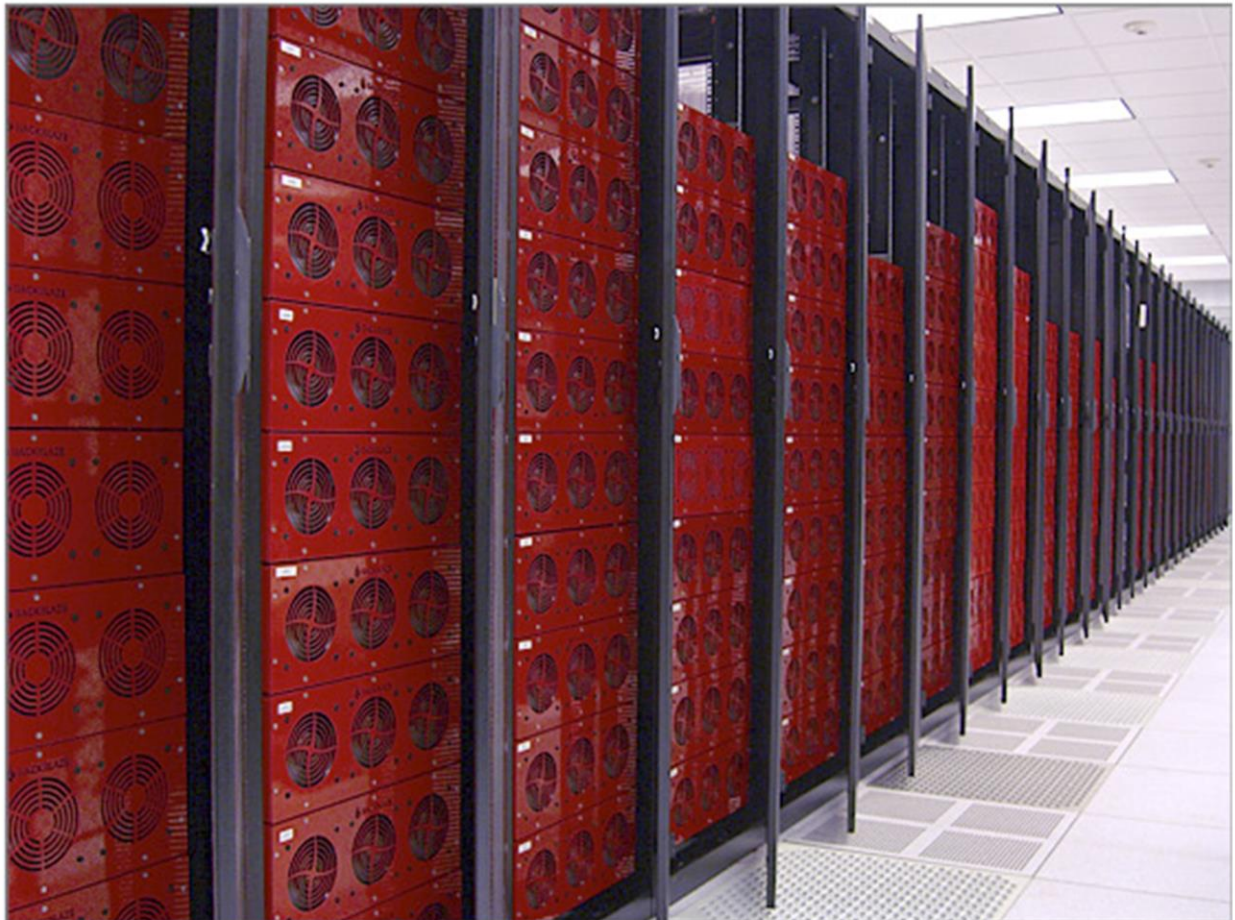
That guy is Sean Harris, and once every two weeks he deploys six pods during an eight-hour work day. Another employee (sometimes it is the CEO) helps Sean lift the 143-pound pods into place. He does the rest. Sean also spends one day each week replacing drives that have gone bad. As of this week, Backblaze has more than 9,000 hard drives in the datacenter, the oldest was purchased four years ago.

“I’ve worn a lot of hats in the tech world starting off in technical support, working in the field as a data warehouse consultant, DBA, and QA Engineer,” says Harris in an email to VentureBeat. “So working in the heart of the operations is both challenging and fun. But my Christmas wish to the company is another IT person... or two.”

All told, Sean Harris replaces approximately 10 drives per week.

In case you were curious, the company's favorite drive is the Hitachi 3TB drive (Hitachi Deskstar 5K3000 HDS5C3030ALA630) because of its low power demand and reliability. Their least favorite is the Western Digital WD20EADS and the Seagate ST32000542AS.

Now go ahead and build your own storage pod facility. Please send us a picture when it's ready!



Backblaze's 500-petabyte data center is more proof of the digital data explosion



Derrick Harris Feb 3, 2014 - 7:52 AM CDT


Cloud backup provider Backblaze has [moved into a new data center in Sacramento](#) capable of storing 500 petabytes, or half an exabyte, of data. It's not full yet (the company [was storing 75 petabytes](#) as of November), but the pace is picking up and it probably will be sooner than some might expect. The crazy part is that Backblaze isn't even that big a company or that widely used a service. Facebook alone is building enough capacity to house 3 exabytes of data in each of its 3 cold storage facilities. Sometimes, I can't help but think that we're just digitally hoarding.



Why hardware vendors need to sell Google and Facebook lousier products if they want to survive



Matt Weinberger  

Mar. 2, 2016, 8:08 PM  8,103

Last month, Google's VP of Infrastructure made an unusual request to computer hardware manufacturers: build hard drives that are less reliable.

It seems silly — why would Google want its [data centers](#) to use drives that are more likely to *lose* data?



Backblaze CEO and co-founder Gleb Budman

Backblaze/YouTube

The answer is simple: Thanks to a radical rethinking in how data centers are built, tech companies like Google don't need a reliable hard drive, as much as they need lots of cheap, high-capacity ones. After all, with Google's massive reach, the data is copied all over the globe anyway.

"[We] already have to have that data somewhere else anyway," writes Google's Eric Brewer in a [blog entry](#). "It's not that we want the disk to lose data, but rather that we can better focus the cost and effort spent trying to avoid data loss for other gains such as capacity or system performance."

Google may find that convincing hardware vendors to heed its advice and play along isn't so easy.

Although the needs of Google and other big customers have changed in recent years, the hardware industry still builds products as if it were 1999.

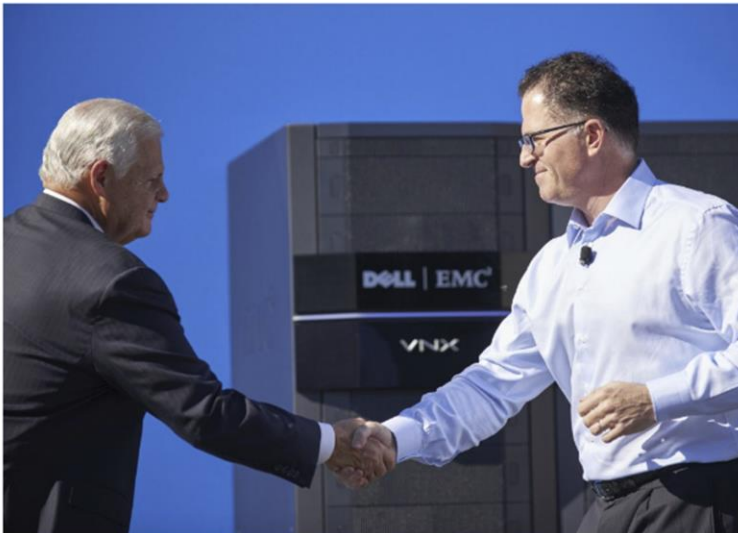
Business Insider talked to Gleb Budman, CEO of popular data storage and backup startup [Backblaze](#), about the market forces that have transformed the industry, and about why big hardware companies, like [Dell](#), HP Enterprise, and IBM will need to do a lot of soul-searching if they want to stay relevant in a world ruled by cloud computing.

"There's not as much need for companies to build big, expensive hardware systems," Budman explains.

'They make billions by not doing it'

Technologically speaking, there's no reason why the established legacy hardware manufacturers couldn't do as Google's Brewer asks, Budman says. But for companies like Dell and HP, doing as Google asks would basically require jettisoning some of their most lucrative businesses.

"They make billions of dollars off not doing it," Budman says. "They haven't been forced to make that shift."



EMC CEO Joe Tucci (left) shakes hands with Dell CEO and founder Michael Dell (right) Dell

Historically, companies like Dell and HP have made a mint from selling storage hardware. In the consumer PC market, hardware is low margin. But when they sell to businesses, they can sell even basic stuff like hard drives at a huge markup, citing additional enterprise features and reliability.

In the past, corporate customers, for lack of a better option, accepted that as the cost of doing business. As the PC started to infiltrate the home and office, companies had to build server rooms and data centers.

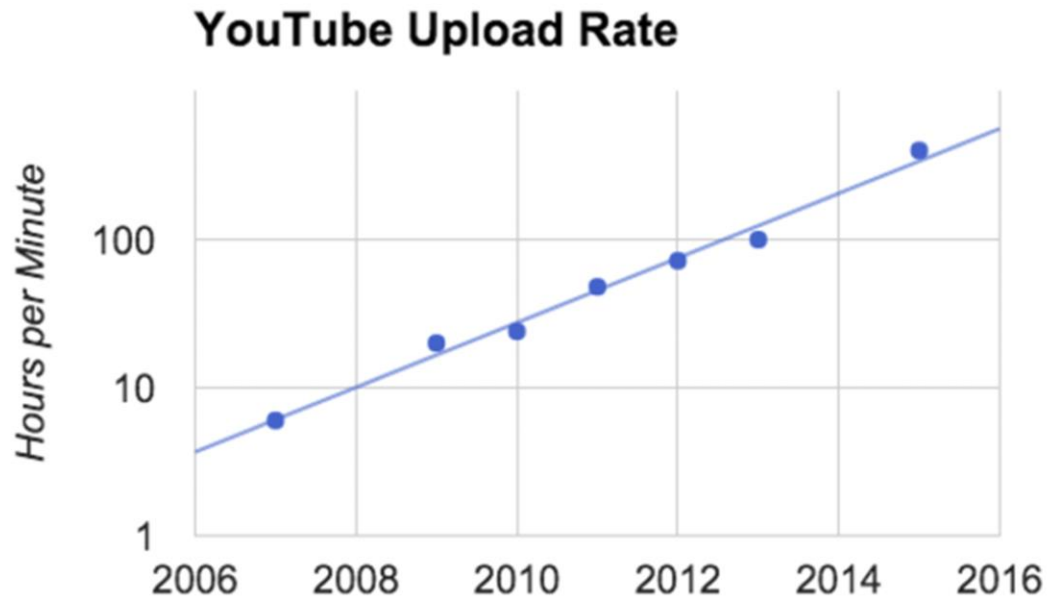
For years, that was how it had to be. But then in the late 2000s, everything changed.

Enter the iPhone

In 2007, Apple introduced the iPhone, kickstarting a whole new era in computing. Suddenly, everybody was carrying an Internet-connected device with them at all times, always. That created an avalanche of new data, all of which needs to be stored somewhere.

When you upload a photo to Instagram, that data has to go somewhere. Similarly, when you get an e-mail on Gmail, that message lives on a server somewhere. And as phones get better, Internet speeds get faster, and services like YouTube and Facebook video get more popular, we're uploading more stuff than ever before.

Check out this chart Brewer shared of the YouTube upload rate, projected through 2016:



Google

This is where [cloud computing platforms](#) like [Amazon Web Services](#), [Google Cloud Platform](#), and [Microsoft Azure](#) fit in. The cloud providers basically store all the videos, photos and other data that consumers and businesses create.

It's incredibly cost-prohibitive for cloud providers to buy pricey enterprise storage products in the quantities they would need to keep up with demand.

"A cloud-first data center is different," Budman says.

Rethinking everything

The bad news for the old guard of hardware vendors is that the cloud providers and internet companies are signaling that they would [vastly prefer not to buy from them](#) if they don't have to.

The best example: Facebook has its [Open Compute Project](#), a multi-company initiative to build smarter data centers that rely less on expensive vendor products and more on intelligently rethinking the basic components — right down to the hard drive.

Earlier in 2016, Facebook opened up its sixth global data center powered entirely by [Open Compute Project](#) tech. It didn't use any equipment from established enterprise vendors. If they *had*, at Facebook's scale, it would have been a massive payday for the vendor of choice.

Amazon Web Services and Google are more secretive about their operations, but they're similarly working hard to reduce their spending on expensive enterprise technology — why spend the money when they can build something in-house that's better and cheaper?



A Facebook engineer slots in a Open Compute Project server Facebook

As the cloud gets more popular, it eats away at spending on data centers, Budman says.

And as that happens, it slowly but surely eats away at the bottom lines of enterprise equipment manufacturers like the Dell-EMC combo. Worse, the cloud vendors aren't buying enough of what they're selling to make up the difference.

That puts the hardware makers in a tough spot.

Sure, Dell, HP, EMC, IBM, and other vendors could focus their energies on selling stuff that the cloud companies need, rather than trying to protect their stakes a shrinking market, Budman says. But it would require a certain willingness to let go of a high-margin business in search of a new niche in a cloud world.

Still, Budman doesn't think all hope is lost for the legacy vendors.

"They've reinvented themselves numerous times," Budman says.

Hacking hardware isn't just cool — it's also good business

[Derrick Harris](#) Oct 26, 2012 - 7:24 AM CDT

One of the best things about cloud computing — as both a business model and an architectural principle — is that hardware really doesn't matter. By and large, as long as applications and systems management software are intelligent enough to run the show, servers, switches and hard drives just need to show up with minimal competency and stay out of the way. If you don't believe me, just ask Backblaze ... or VMware (s vmw) ... or Facebook (s fb).

Smart companies trying to deliver services over the web realize that they're not in the business of pleasing a CIO, but of pleasing consumers. And consumers don't care what's under the hood as long as the service works and their lives aren't interrupted by a downed server.

That's kind of the reason that cloud computing exists, and has become such a successful delivery model for IT resources. Users get generic server specs on virtual machines, but all the magic happens at the layers above. Smart developers can take advantage of features such as auto-scaling and failover, as well as myriad open source components and open(ish) APIs, to piece together applications that might not look pretty, but stay online and don't cost a fortune to run.

The cloud underpins a lot of applications that many web users couldn't live without, [from Instagram](#) (s fb) [to Netflix](#) (s nflx). They all required some architectural creativity along the way to grow into what they've become, and the cloud enabled that.

What's good for the goose ...

Life shouldn't be any different just because a company decides to run its own servers rather than rent ephemeral boxes from Amazon Web Services (s amzn). Especially for large-scale web applications or services, it might make good business sense to eschew the traditionally expensive world of off-the-shelf computing hardware and just build the cheapest-possible gear that gets the job done — gear that ends up looking a lot like those generic cloud computing resources.

Google wrote the book on this by designing its own servers, data centers and, [it appears, networking switches](#), and the company has done alright for itself. Facebook, too, is lauded for its custom-built hardware and data centers. Both companies have gotten so good at designing gear and sourcing components that they [don't necessarily need buy much of anything](#) from mainstream vendors in order to fill their data centers.



Facebook's custom-built battery cabinet

But neither company would consider itself a hardware company (which is why Facebook was willing to [open source its designs under the Open Compute Project banner](#)), they just realize that a little hacking can go a long way toward delivering a better service. Economically, custom gear that does away with extraneous bells and whistles while adding performance where needed means a lower sticker price, a lower power bill and a better user experience. Whatever reliability is lost by removing fans, server cases and vendor software is made up for by smart software engineers who design systems that expect gear to fail, [and to keep running when that happens](#).

Google is so confident in its software it [promises zero planned downtime for Gmail](#) and achieves higher than 99 percent uptime for the service overall. And ask yourself the last time you remember Facebook, with its 950 million users, crashing. In the enterprise IT world, it's this type of infrastructural intelligence that's driving [the software-defined network movement](#), and [VMware's vision of software-defined data centers](#).

Backblaze blowback

This is why I was a little shocked to see some of the negative comments when I wrote recently about cloud-storage startup Backblaze's efforts to deal with last year's hard drive shortage by [sourcing consumer-grade external hard drives from Costcos](#) around the country. Given the choices — back off its unlimited storage for \$5 promise, or find a way to procure capacity on the cheap — Backblaze almost certainly made the right choice. The software that runs the service expects hard drives to fail, and backing up data doesn't require blazing fast data access. As long as consumer drives aren't crashing by the hundreds, users don't notice a thing.

What's more, that epic feat of hackery wouldn't have been possible if not for Backblaze's even bigger contribution to cloud-service design — [its \\$7,400 135TB storage pods](#). Because it designs and builds its own infrastructure, the company had the luxury of swapping out the most critical component without worrying about voiding a warranty or messing up something in a storage array whose blueprints it hadn't seen. It's possible that the then-bootstrapped Backblaze wouldn't have made it out of the hard drive shortage alive, or at least without some upset customers, had it not been so in tune with its hardware needs.



BackBlaze's storage pods in a data center

You can do it, too!

Given all this, it's not insignificant that Facebook on Wednesday [shared some tips to deploying its Open Compute servers in co-location facilities](#) that might not be designed to handle custom rack designs. After all, unless you're building your data centers like Facebook, Google [or eBay](#) ([s ebay](#)), you have to play by your colo provider's rules. Guidance from Facebook on actually deploying the servers under real-world circumstances makes Open Compute less good in theory and more good in practice.

Traditional big companies might not line up to deploy Open Compute gear as is — they'll understandably wait until server makers such as Dell ([s dell](#)) and HP ([s hpq](#)) productize the designs. But companies that want to be big like Facebook or Google (or even in sheer capacity a la Backblaze) ought to pay attention. Facebook's server designs could be a great starting point from which to build your own specialized gear to make sure your application and your users' experience are never at the mercy of fate or a vendor's bottom line.

Feature image courtesy of [Shutterstock user Jason Winter](#).