


-  Blog Home
-  Feed RSS

The developers of Backblaze post on this weblog about the product, keeping data safe and other rants about life. Visit the How it Works page to learn more.

**Start backing up your files online:** [Get started](#)

[Download Now](#) 

- Tags
- [Backup Awareness Month](#)
  - [Mac Love](#)
  - [Offers](#)
  - [Release](#)
  - [Uncategorized](#)

-  Google
-  MY YAHOO!
-  Bloglines

# Petabytes on a budget: How to build cheap cloud storage

Tim Nufire September 1

**MEET THE BACKBLAZE POD**

67 TERABYTES FOR \$7,867



At Backblaze, we provide unlimited storage to our customers for only \$5 per month, so we had to figure out how to store hundreds of petabytes of customer data in a reliable, scalable way—and keep our costs low. After looking at several overpriced commercial solutions, we decided to build our own custom Backblaze Storage Pods: 67 terabyte 4U servers for \$7,867.

In this post, we'll share how to make one of these storage pods, and you're welcome to use this design. Our hope is that by sharing, others can benefit and, ultimately, refine this concept and send improvements back to us. Evolving and lowering costs is critical to our continuing success at Backblaze.

Below is a video that shows a 3-D model of the Backblaze Storage Pod. Continue reading to learn the exact details of the design.



You can download the full 3-D model of the Backblaze Storage Pod here.

## Backblaze Needs Plenty of Reliable, Cheap Storage

To say that Backblaze needs lots of storage is an understatement. We're a backup service, so our datacenter contains a complete copy of all of our customers' data, plus multiple versions of files that change. In rough terms, every time one of our customers buys a hard drive, Backblaze needs another hard drive. A long time ago we stopped measuring storage in our datacenter in gigabytes or terabytes and started measuring in petabytes.

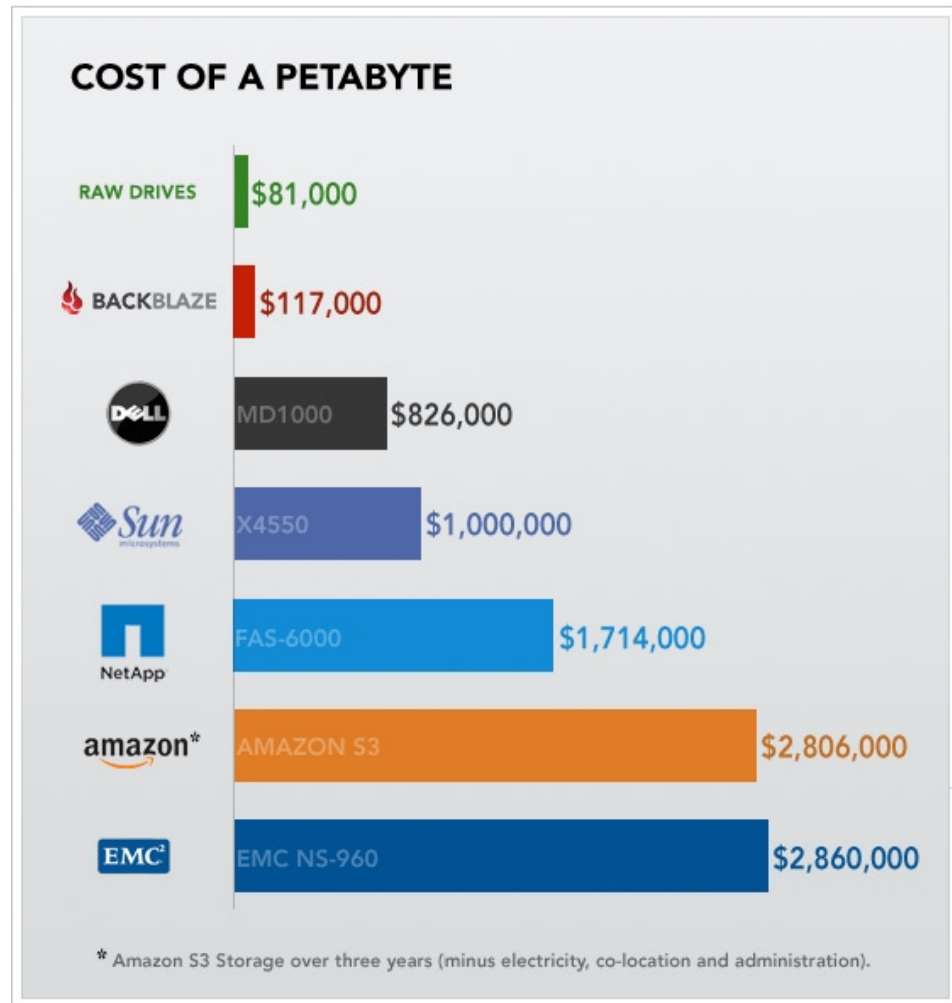
To get a sense of what this looks like, here is a shot of me deploying new pods in our datacenter. The small stack of six pods in the rack I'm working on contains just under half a petabyte of storage.



To offer our service at a reasonable price, we need affordable storage at a multi-petabyte scale.

### No One Sells Cheap Storage, so We Designed It

Before realizing that we had to solve this storage problem ourselves, we considered Amazon S3, Dell or Sun Servers, NetApp Filers, EMC SAN, etc. As we investigated these traditional off-the-shelf solutions, we became increasingly disillusioned by the expense. When you strip away the marketing terms and fancy logos from any storage solution, data ends up on a hard drive. But when we priced various off-the-shelf solutions, the cost was 10 times as much (or more) than the raw hard drives. Here's a comparison chart of the price for one petabyte from various vendors:

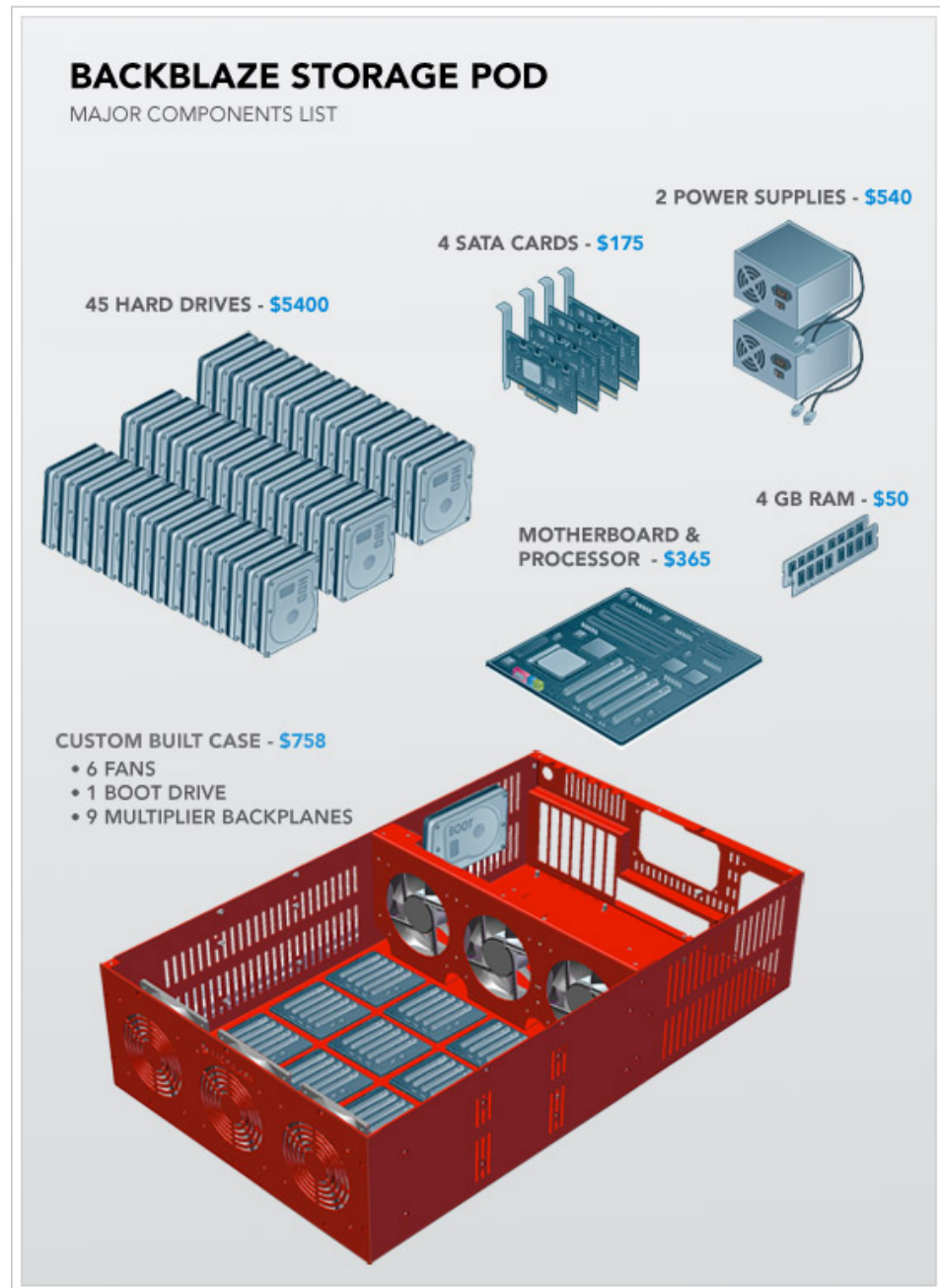


Based on the expense, we decided to build our own Backblaze Storage Pods. We had two primary goals: Keep upfront costs low by using consumer-grade drives and readily available commodity components and be as power and space efficient as possible by using green components and squeezing a lot of storage into a small box.

The result is a 4U rack-mounted Linux-based server that contains 67 terabytes at a material cost of \$7,867, the bulk of which goes to purchase the drives themselves. This translates to just three-tenths of one penny per gigabyte per month over the course of three years. Even including the surrounding costs—such as electricity, bandwidth, space rental, and IT administrators' salaries—Backblaze spends one-tenth of the price in comparison to using Amazon S3, Dell Servers, NetApp Filers, or an EMC SAN.

## What Makes a Backblaze Storage Pod

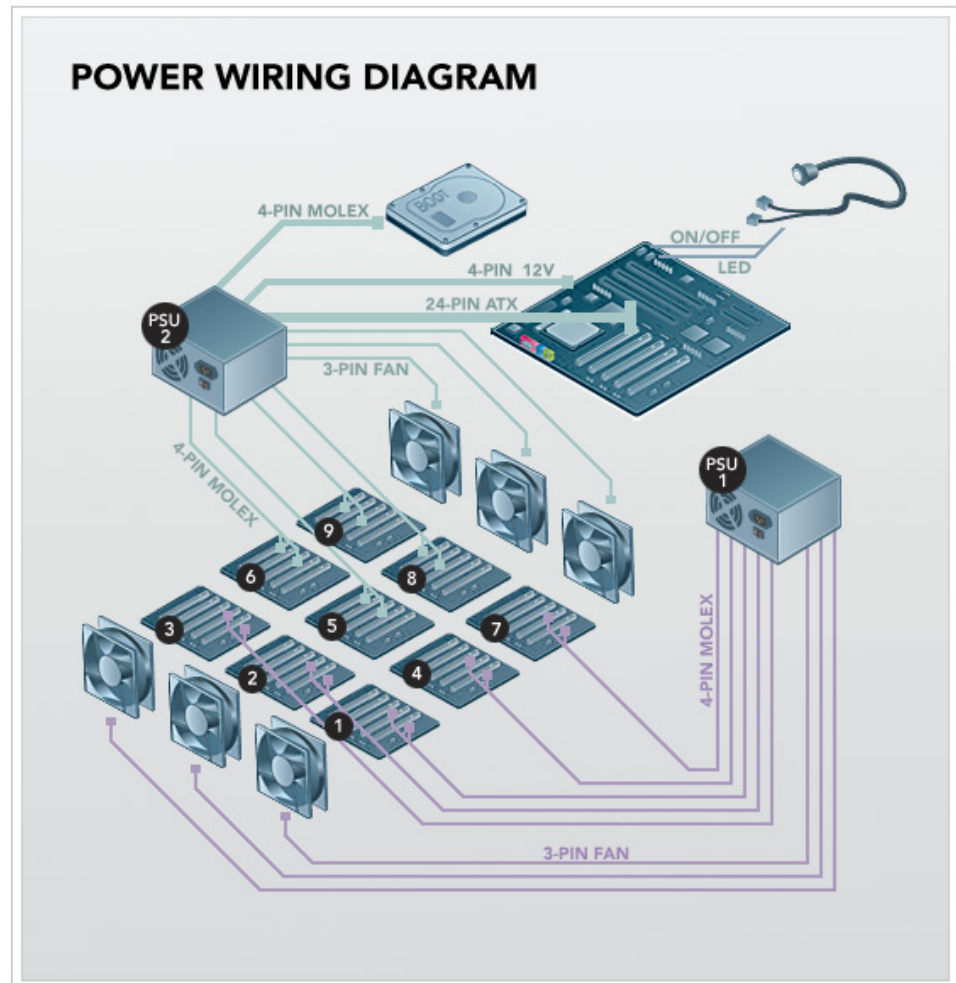
A Backblaze Storage Pod is a self-contained unit that puts storage online. It's made up of a custom metal case with commodity hardware inside. Specifically, one pod contains one Intel Motherboard with four SATA cards plugged into it. The nine SATA cables run from the cards to nine port multiplier backplanes that each have five hard drives plugged directly into them (45 hard drives in total).



Above is an exploded diagram, and you can see a detailed parts list in Appendix A at the bottom of this post. The two most important factors to note are that the cost of the hard drives dominates the price of the overall pod and that the rest of the system is made entirely of commodity parts.

## Wiring It Up: How to Assemble a Backblaze Storage Pod

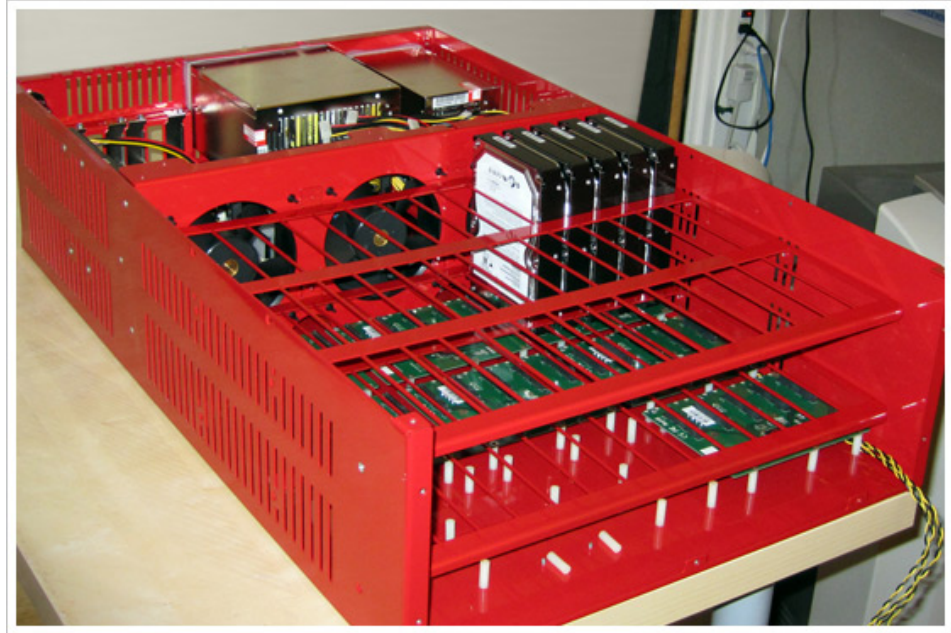
The power wiring diagram of a Backblaze Storage Pod is seen below. Power supply units (PSUs) provide most of their power on two different voltages: 5V and 12V. We use two power supplies in the pod because 45 drives draw a lot of 5V power, yet high wattage ATX PSUs provide most of their power on 12V. This is not an accident: 1,500 watt and larger ATX power supplies are designed for powerful 3-D graphics cards that need the extra power on the 12V rail. We could have switched to a power supply designed for servers, but two ATX PSUs are cheaper.



PSU1 powers the front three fans and port multiplier backplanes 1,2,3,4, and 7. PSU2 powers everything else. (See Appendix A for a detailed list of the custom connectors on each PSU.) To power the port multiplier backplanes, the power cables run from the PSUs through four holes in the divider metal plate that holds the fans at the center of the case (near the base of the fans) and then continue to the underside of the nine backplanes. Each port multiplier backplane has two molex male connectors on the underside. Hard drives draw the most power during initial spin-up, so if you power up both PSUs at the same time, it can draw a large (14 amp) spike of 120V power from the socket. We recommend powering up PSU1 first, waiting until the drives are spun-up (and the power draw decreases to a reasonable level), and then powering up PSU2. Fully booted, the entire pod will draw approximately 4.8 amps idle and up to 5.6 amps under heavy load.

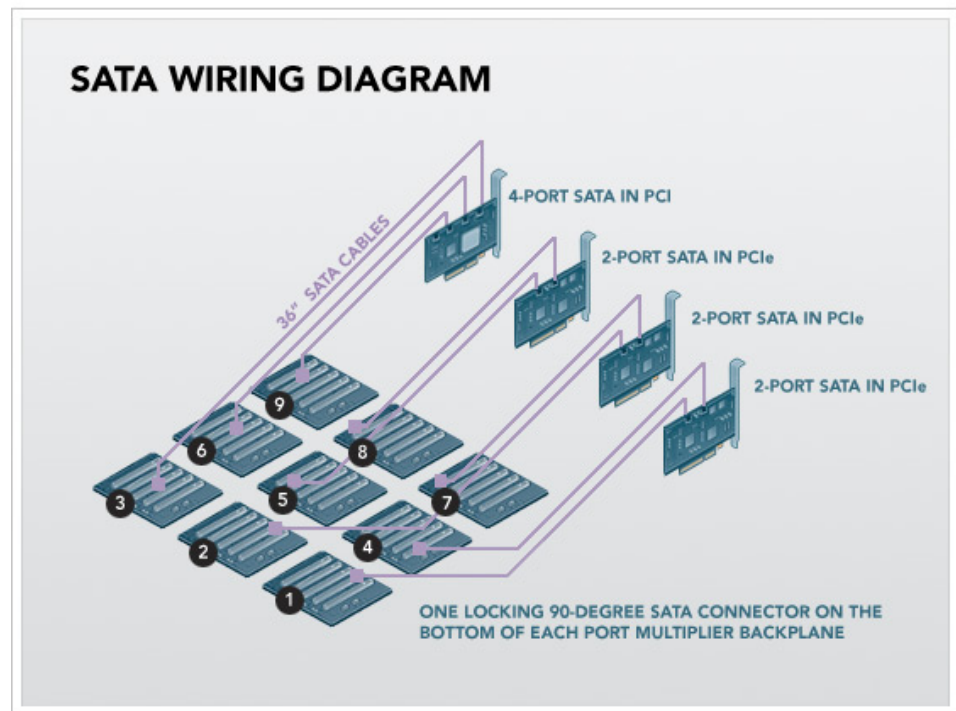
Below is a picture of a partially assembled Backblaze Storage Pod (click on the photo for a larger image). The metal case has screws mounted on the bottom, facing upward, where we attach nylon standoffs (the small white pieces in the picture below). Nylon helps dampen vibration, and this dampening is a critical aspect of server design. The

circuit boards shown on top of the nylon standoffs are a few of the nine SATA port multiplier backplanes that take a single SATA connection on their underside and allow five hard drives to be mounted vertically and plugged into the topside of the board. All the power and SATA cables run underneath the port multiplier backplanes. One of the backplanes in the picture below is fully populated with hard drives to show the positioning.



A note about drive vibration: The drives vibrate too much if you leave them sitting as shown in the picture above, so we add an “anti-vibration sleeve” (essentially a rubber band) around the hard drive in between the red metal grid and the drives. This seats the drives tightly in the rubber. We also lay a large (16” x 17” x 1/8”) piece of foam along top of the hard drives after all 45 are in the case. The lid then screws down on top of the foam to hold the drives securely. In the future, we will dedicate an entire blog post to vibration.

The SATA wiring diagram is seen below.

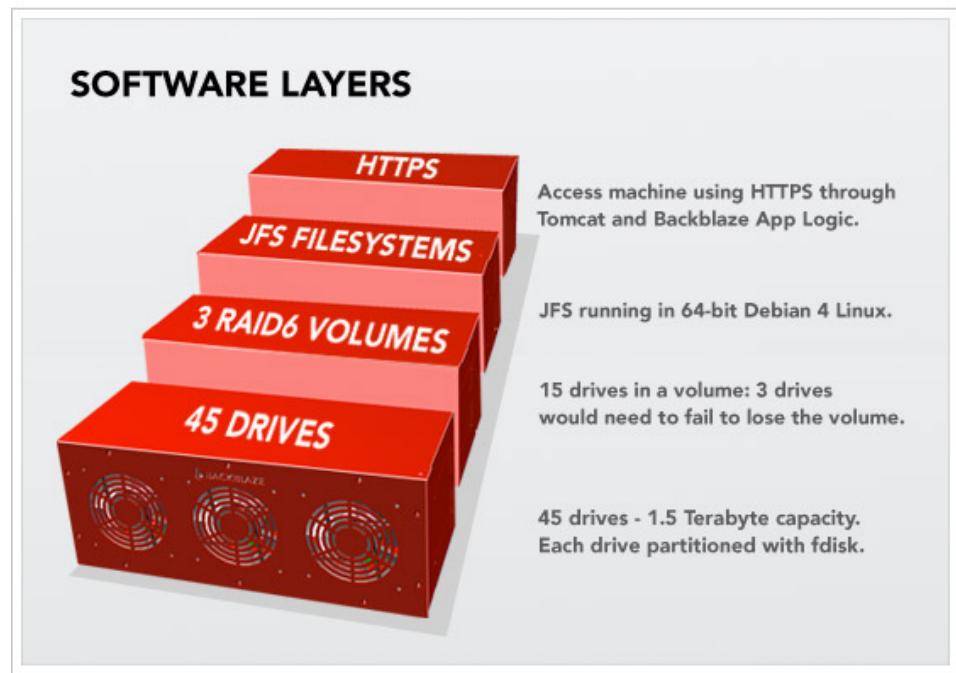


The Intel Motherboard has four SATA cards plugged into it: three SYBA two-port SATA cards and one Addonics four-port card. The nine SATA cables connect to the top of the SATA cards and run in tandem with the power cables. All nine SATA cables measure 36 inches and use locking 90-degree connectors on the backplane end and non-locking straight connectors into the SATA cards.

A note about SATA chipsets: Each of the port multiplier backplanes has a Silicon Image Sil3726 chip so that five drives can be attached to one SATA port. Each of the SYBA two-port PCIe SATA cards has a Silicon Image Sil3132, and the four-port PCI Addonics card has a Silicon Image Sil3124 chip. We use only three of the four available ports on the Addonics card because we have only nine backplanes. We don't use the SATA ports on the motherboard because, despite Intel's claims of port multiplier support in their ICH10 south bridge, we noticed strange results in our performance tests. Silicon Image pioneered port multiplier technology, and their chips work best together.

## A Backblaze Storage Pod Runs Free Software

A Backblaze Storage Pod isn't a complete building block until it boots and is on the network. The pods boot 64-bit Debian 4 Linux and the JFS file system, and they are self-contained appliances, where all access to and from the pods is through HTTPS. Below is a layer cake diagram.



Starting at the bottom, there are 45 hard drives exposed through the SATA controllers. We then use the fdisk tool on Linux to create one partition per drive. On top of that, we cluster 15 hard drives into a single RAID6 volume with two parity drives (out of the 15). The RAID6 is created with the mdadm utility. On top of that is the JFS file system, and the only access we then allow to this totally self-contained storage building block is through HTTPS running custom Backblaze application layer logic in Apache Tomcat 5.5. After taking all this into account, the formatted (useable) space is 87 percent of the raw hard drive totals. One of the most important concepts here is that to store or retrieve data with a Backblaze Storage Pod, it is always through HTTPS. There is no iSCSI, no NFS, no SQL, no Fibre Channel. None of those technologies scales as cheaply, reliably, goes as big, nor can be managed as easily as stand-alone pods with their own IP address waiting for requests on HTTPS.

### A Backblaze Storage Pod is a Building Block

We have been extremely happy with the reliability and excellent performance of the pods, and a Backblaze Storage Pod is a fully contained storage server. But the intelligence of where to store data and how to encrypt it, deduplicate it, and index it is all at a higher level (outside the scope of this blog post). When you run a datacenter with thousands of hard drives, CPUs, motherboards, and power supplies, you are going to have hardware failures—it's irrefutable. Backblaze Storage Pods are building blocks upon which a larger system can be organized that doesn't allow for a single point of failure. Each pod in itself is just a big chunk of raw storage for an inexpensive price; it is not a "solution" in itself.

### Cloud Storage: The Next Step

The first step to building cheap cloud storage is to already have cheap storage, and above we demonstrate how to create your own. If all you need is cheap storage, this may suffice. If you need to build a cloud, you've got more work ahead of you.

Building a cloud includes not only deploying a large quantity of hardware, but, critically, deploying software to manage it. At Backblaze we have developed software that de-duplicates and chops data into blocks; encrypts and transfers it for backup; reassembles, decrypts, re-duplicates, and packages the data for recovery; and



monitors and manages the entire cloud storage system. This process is proprietary technology that we have developed over the years.

You may have your own system for this process and incorporate the Backblaze Storage Pod design, or you may simply seek inexpensive storage that won't be deployed as part of a cloud. In either case, you're free to use the storage pod design above. If you do, we would appreciate credit at Backblaze and welcome any insights, though this isn't a requirement. Please note that because we're not selling the design or the storage pods themselves, we provide no support nor warranties.

Coming next: In the next few weeks, we'll talk about iPhone vibration sensors, swiss cheese pod designs, why electricity costs more than bandwidth, and more about the design of big cloud storage.

## Credits and Standing on the Shoulders of Giants

The Backblaze Storage Pod design would not have been possible without an enormous amount of help, usually requested with little notice, from some amazingly smart and generous people who answered our questions, worked with us, and provided key insights at critical moments. First, we thank Chris Robertson for the inspiration to build our own storage and his early work on prototypes; Kurt Shafer for advice on metal work and the concept of "furniture" for circuit boards; Dominic Giampaolo from Apple Computer for his advice on hard drives, vibration, and certifications; Stuart Cheshire from Apple Computer and Nick Tingle from Alcatel-Lucent for low-level network advice; Aaron Emigh (EVP & GM, Core Technology) at Six Apart for his help on initial design work; Gary Orenstein for insight into drive reliability and the storage industry in general; Jonathan Beck for invaluable advice on vibration, fans, cooling, and case design; Steve Smith (Senior Design Manager), Imran Pasha (Director of Software Engineering), and Alex Chervet (Director of Strategic Marketing) at Silicon Image who helped us debug SATA protocol problems and loaned us 10 different SATA cards to test against; James Lee at Chyang Fun Industries in Taiwan for customizing SATA boards to simplify our design; Wes Slimick, Richard Crockett, Don Shields, and Robert Knowles from Western Digital for their help debugging Western Digital drive logs; Christa Carey, Jennifer Hurd, and Shirley Evely at Protocase for putting up with hundreds of small 3-D case design tweaks; Chester Yeung at Central Computer for coming through quickly and repeatedly with locally supplied parts when it really mattered; Mason Lee at Zippy for power supply advice and custom wiring harnesses; and Angela Lai for knowing just the right people and providing gracious introductions.

Finally, we thank the thousands of engineers who slaved away for millions of hours to bring us the pod components that are either inexpensive or totally free, such as the Intel Processor, Gigabit Ethernet, ridiculously dense hard drives, Linux, Tomcat, JFS, etc. We realize we're standing on the shoulders of giants.

## Appendix A: Detailed Backblaze Storage Pod Parts List

Item	Qty	Price	Total
<b>1.5 TB SATA Data Drive</b> Seagate ST31500341AS 1.5TB Barracuda 7200.11 SATA 3Gb/s 3.5"	45	\$120.00	\$5,400
<b>4U Enclosure</b> Custom Designed 4U Red Backblaze Storage Pod Enclosure	1	\$748.00	\$748
<b>760 Watt Power Supply</b> Zippy PSM-5760 760 Watt Power Supply with Custom Wiring (see below)	2	\$270.00	\$540
<b>Port Multiplier Backplanes</b> Chyang Fun Industry (CFI Group) CFI-B53PM 5 Port Backplane (Sil3726)	9	\$42.00	\$378
<b>3.3 GHz Intel Core 2 CPU</b> Intel E8600 Wolfdale 3.33 GHz LGA 775 65W Dual-Core Processor	1	\$280.00	\$280

<b>2 Port PCIe SATA II Card</b> Syba SD-SA2PEX-2IR PCI Express SATA II Controller Card (Sil3132)	3	\$35.00	\$105
<b>4 Port PCI SATA II Card</b> Addonics ADSA4R5 4-Port SATA II PCI Controller (Sil3124)	1	\$70.00	\$70
<b>Motherboard</b> Intel BOXDG43NB LGA 775 G43 ATX Motherboard	1	\$85.00	\$85
<b>Case Fan</b> Mechatronics G1238M12B1-FSR 120 x 38 mm 2,800 RPM 12V Fan	6	\$13.60	\$82
<b>4GB DDR2 800 RAM</b> Kingston KVR800D2N6K2/4G 2x2GB 240-Pin SDRAM DDR2 800 (PC2 6400)	1	\$50.00	\$50
<b>80 GB PATA Boot Drive</b> Western Digital Caviar WD800BB 80GB 7200 RPM IDE Ultra ATA100 3.5"	1	\$38.00	\$38
<b>On/Off Switch</b> FrozenCPU ele-302 Bulgin Vandal Momentary LED Power Switch 12" 2-pin	1	\$30.00	\$30
<b>SATA II Cable</b> SATA II Cable, 90 Degrees/straight with Locking Connectors	9	\$2.00	\$18
<b>Nylon Backplane Standoffs</b> Fastener SuperStore 1/4" Round Nylon Standoffs Female/Female 4-40 x 3/4"	72	\$0.17	\$12
<b>HD Anti-Vibration Sleeves</b> Aero Rubber Co. 3.5 x .500 inch EPDM (0.03" Wall)	45	\$0.23	\$10
<b>Power Supply Vibration Dampener</b> Vantec VDK-PSU Power Supply Vibration Dampener	2	\$4.50	\$9
<b>Fan Mount (front)</b> Acousti Ultra Soft Anti-Vibration Fan Mount AFM02	12	\$0.40	\$5
<b>Fan Mount (middle)</b> Acousti Ultra Soft Anti-Vibration Fan Mount AFM03	12	\$0.40	\$5
<b>Nylon Screws</b> Small Parts MPN-0440-06P-C Nylon Pan Head Phillips Screw 4-40 x 3/8"	72	\$0.02	\$1
<b>Foam Rubber Pad</b> House of Foam 16" x 17" x 1/8" Foam Rubber Pad	1	\$1.00	\$1
<b>TOTAL</b>			<b>\$7,867</b>

**Custom wiring harnesses for PSU1 (1st Zippy power supply):**

- 5x 4-pin 90-degree molex HD connectors with two connectors each. Length should be 36" to the farthest connector, 32.5" to the closest (3.5" apart)
- 3x 4-pin 12V fan connectors that should be 32" in length with extender and RPM signal that can attach to motherboard

**Custom wiring harnesses for PSU2 (2nd Zippy power supply):**

- 1x 24-pin motherboard connector, 8"
- 1x 4-pin ATX12V for CPU, 8"
- 4x 4-pin 90-degree molex HD connectors with two connectors each. Length should be 36" to the farthest connector, 32.5" to the closest (3.5" apart)
- 1x 4-pin 90-degree molex connector, 12" long

- 3x 3-pin, 12V fan connectors, 12" long, with extender for RPM signal that can attach to motherboard

### SATA Chipsets

- Sil3726 on each port multiplier backplane to attach five drives to one SATA port.
- Sil3132 on each of the three PCIe SATA cards to attach two backplanes each (six ports total)
- Sil3124 on the one PCI SATA card to attach up to four port multiplier backplanes (we only use three of the four ports)

Share this with: These icons link to social bookmarking sites where readers can share and discover new web pages.

- 
- 
- 

Tags: [TechBytes](#), [Cloud Storage](#)