

STORAGE DEVELOPER CONFERENCE



Fremont, CA
September 12-15, 2022

BY Developers FOR Developers

A **SNIA** Event

More Than Just a Bucket of Bits

Cloud Object Storage turns Sweet Sixteen

Pat Patterson

Chief Technical Evangelist, Backblaze

@metadaddy / pat@backblaze.com

Press release

Amazon Web Services Launches

March 14, 2006 at 3:00 AM EST

S3 Provides Application Programming Interface for Highly Scalable Reliable, Low-Latency Storage at Very Low Costs

SEATTLE--(BUSINESS WIRE)--March 14, 2006-- Amazon Web Services today announced "Amazon S3(TM)," a simple storage service that offers software developers a highly scalable, reliable, and low-latency data storage infrastructure at very low costs. Amazon S3 is available today at <http://aws.amazon.com/s3>.

Amazon S3 is storage for the Internet. It's designed to make web-scale computing easier for developers. Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web. It gives any developer access to the same highly scalable, reliable, fast, inexpensive data storage infrastructure that Amazon uses to run its own global network of web sites. The service aims to maximize benefits of scale and to pass those benefits on to developers.

Amazon S3 Functionality

Amazon S3 is intentionally built with a minimal feature set. The focus is on simplicity and robustness.

- Write, read, and delete **objects** containing from 1 byte to 5 gigabytes of data each. The number of objects that can be stored is unlimited.
- Each object is stored and retrieved via a unique developer-assigned **key**.
- Objects can be made private or public, and rights can be assigned to specific users.
- Uses standards-based **REST** and **SOAP** interfaces designed to work with any **Internet-development toolkit**.

The POSIX File System API

```
mkdir("/mydirectory", S_IRWXU);  
int fd = creat("/mydirectory/myfile", S_IRWXU);  
write(fd, buffer, count);  
close(fd);
```

The operating system must manage a file descriptor for **every** file being accessed by **every** process.

The S3 API

```
PUT /mybucket HTTP/1.1
```

```
Host: s3.amazonaws.com
```

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<CreateBucketConfiguration>
```

```
  <LocationConstraint>us-west-1</LocationConstraint>
```

```
</CreateBucketConfiguration>
```

```
PUT /mybucket/mykey HTTP/1.1
```

```
Host: s3.amazonaws.com
```

Body

The only state is in the buckets and objects in storage.

Cloud Object Semantics

How Does S3 Work?

Amazon isn't telling

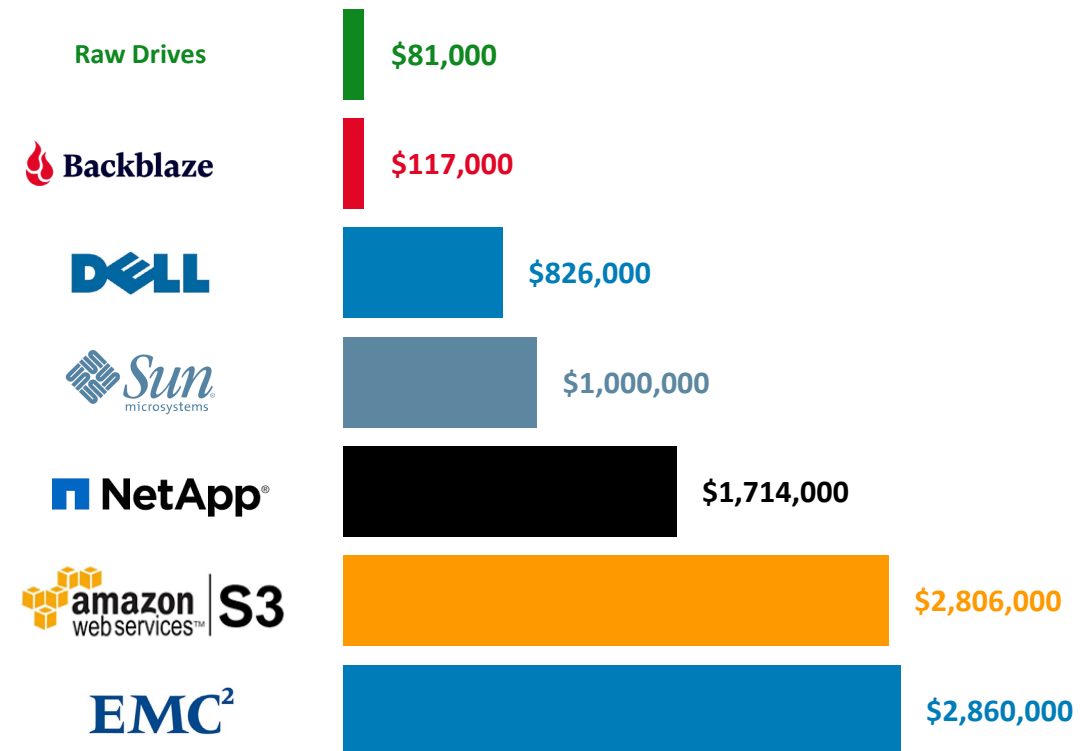
We're not Amazon

Why We Built Backblaze Storage Cloud

We wanted to use Amazon Storage but it was too expensive

1 Petabyte = 1 million GBytes

Cost of a Petabyte
(September 2009)



Scale: Step 1 - Put Drives Online

Connect hard drives to the internet as cheaply as possible.

This is a pod.

Backblaze Storage Pod
1 server, 60 drives



What's in a Pod?

- Custom Web Application
- Apache Tomcat
- Debian
- EXT4 File System

- 60 x 16 TB drives = 960 TB

- HTTPS - Internet accessible
- No iSCSI/NFS/SQL/Fibre Channel



Build Your Own Pod!

60 x Seagate 16 TB drives	\$17399.40	(retail)
Case, mobo, PSU etc	\$3464.67	
Total	\$20864.07	
Cost/GB	\$0.022	

Backblaze Storage Pod 6.0 Parts List

Part Description	MFG/Supplier	MFG PN	QTY	Unit Cost	Ext Cost	Notes
SCREW, 4-40 X 5/16 PHIL PAN ZPS ROH	MCMMASTER	90272A107	60	0.02	1.20	1
SCREW, 6-32 X 1/4 PHIL PAN ZPS	MCMMASTER	90272A144	12	0.15	1.83	1
SCREW, 6-32 X 1/4 100D PHIL FLAT ZP	MCMMASTER	90471A205	39	0.20	7.76	1
POWER SUPPLY VIBRATION DAMPENERS KIT	Quiet PC	FSPSUC	2	4.45	8.90	
SOFT FAN MOUNT, AFM03B (TWO TAB END)	Quiet PC	AFM03B	12	0.42	4.99	
INTEL XEON PROCESSOR	INTEL	XEON Quad-Core E5-1620 V2	1	343.94	343.94	
FAN AXIAL 120X25MM VAPO 12VDC	SUNON	3161/MEC0251V1-000U-G99	3	10.60	31.80	
BACKPLANE, 5-PORT SATA	Sunrich Tech	S-331	12	45.68	548.10	2
SATA CARD, 4-PORT PCI EXPRESS RAID	Sunrich Tech	A-540	3	57.10	171.30	2
8GB MEMORY,DDR3-1600	HYNIX	HMT31GR7CFR4C-PB	4	89.49	357.96	
MOTHER BOARD, X9	Supermicro	MBD-X9SRH-7TF-O	1	539.50	539.50	
CABLE, SATA RA-TO-STR 1M LOCKING	Nippon Labs		12	3.33	39.90	3
CABLE ASSEMBLY, WIRED ON-OFF SWITCH	Primochill	PSW-SMV22-R-R	1	14.95	14.95	
POWER SUPPLY, EVGA NEX750G	EVGA	120-G1-0750-XR	2	119.99	239.98	
PS1 HARNESS W/ 24 PIN CONN	NORTRA		1	33.00	33.00	
PS2 HARNESS W/ 24 PIN CONN	NORTRA		1	31.84	31.84	
PIGTAIL,24 PIN CONN EVGA750G	NORTRA		2	16.43	32.85	
Screw,M3x5MM lg. Phillip head	MCMMASTER	92010A114	4	0.95	3.81	1
Standoff,m3x5mm lg. hex,ss	MCMMASTER	93655A800	4	0.69	2.74	
DYNATRON R13 1U SERVER CPU FAN	DYNATRON	R13 1U SERVER CPU FAN	1	45.71	45.71	
Foam strip, 1/2" Thick, 3/4" Width, 17"	MCMMASTER	8512K96	1	0.55	0.55	
Cable Tie, 8.3 inch L x 0.225 inch	Panduit	SG200S-C	4	0.25	1.00	
Cable Tie, 4" length	MCMMASTER	7130K12	2	0.03	0.06	
Drive Guides			120	0.25	30.00	
Label,Serial-Model,Transducer, Blnk	HN Lockwood		30	0.20	6.00	
Chassis	Lor Van		1	995.00	995.00	
				Total	3,494.67	

Notes

- 1 - Sold in packages of 100, used 100 package price for Extended Cost
- 2 - Sunrich and CFI make the recommended backplanes and Sunrich and Syba make the recommended SATA Cards
- 3 - Nippon Labs makes the recommended SATA cables, but others may work



Scale: Step 2 - Backblaze Vaults

Group 20 pods into a “Backblaze Vault” with software.
Vaults are software.

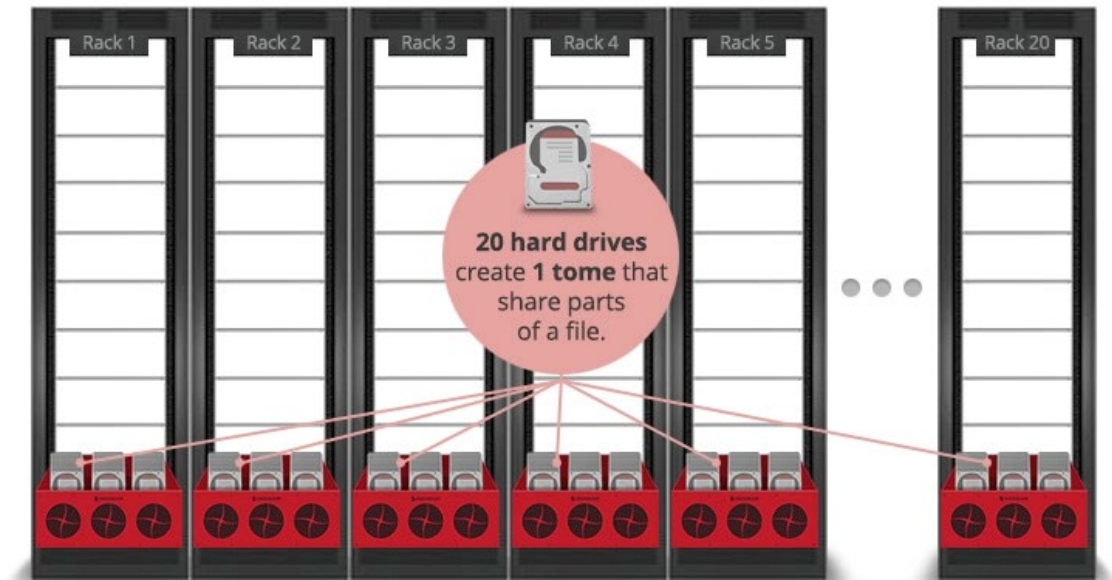
Backblaze Vault - 20 pods

Backblaze Storage Pod - 1 server, 60 drives

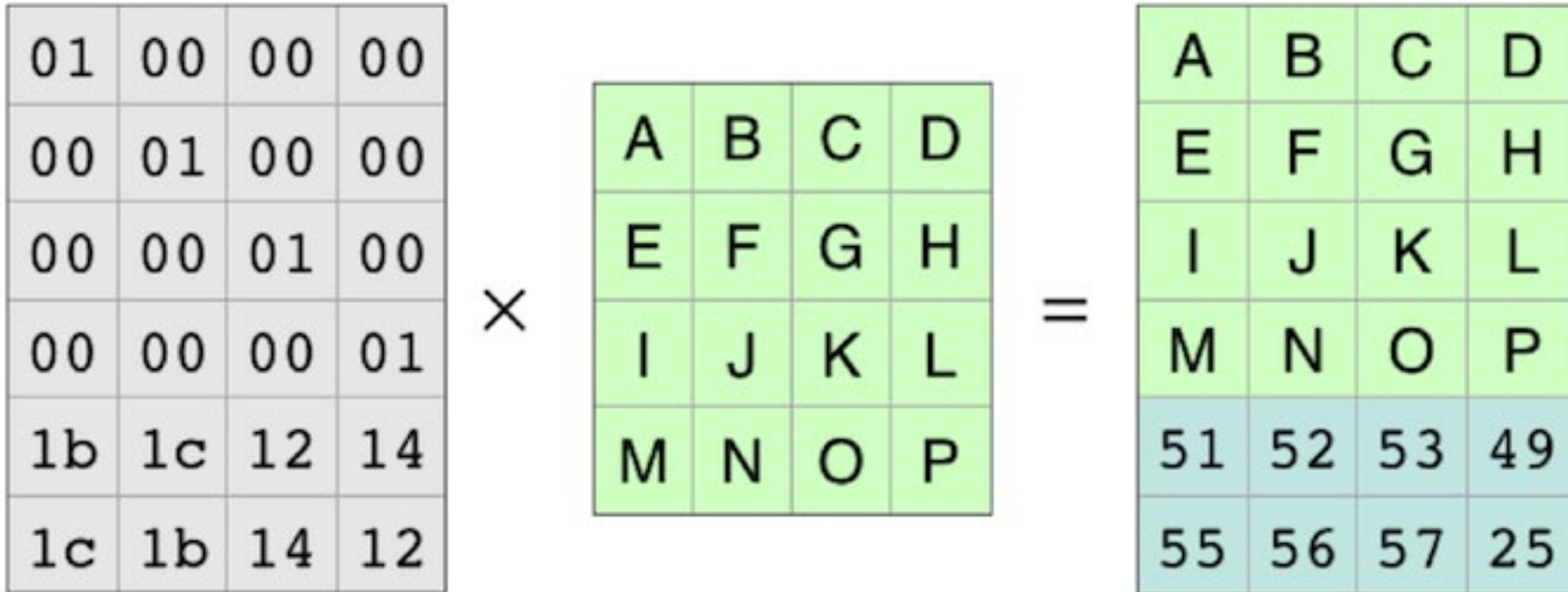


Distributing Data Across 20 Storage Pods

- ✧ Vault = 20 Storage Pods
- ✧ Tome = 20 drives, each in the same position in each Storage Pod
- ✧ Each file is split into 20 shards with Reed-Solomon Erasure Coding
 - ✧ 17 data + 3 parity (< 16 TB drives)
 - ✧ 16 data + 4 parity (\geq 16 TB drives)
- ✧ Shards are written across a single tome
- ✧ 99.999999999% data durability (11 nines)
- ✧ 960 TB/pod
 - x 16 data pods/vault
 - = 15.36 PB/vault



Reed-Solomon Erasure Coding



Scale: Step 3 - Backblaze Clusters

Now group the vaults into a “Backblaze Cluster” with software. Clusters are software.

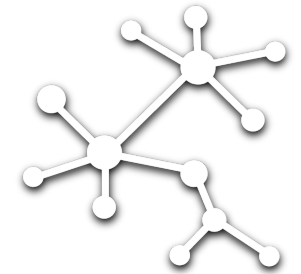
Backblaze Cluster **100 vaults** **2,000 pods** **120,000 hard drives**

Backblaze Vault - 20 pods

Backblaze Storage Pod - 1 server, 60 drives



Scalable
x 100 vaults = 1 cluster



A Cluster is an 'Instance' of Backblaze

- 🔗 One or two clusters per data center
- 🔗 Cluster-wide services
 - ✦ API - Apache Tomcat
 - ✦ Metadata - Apache Cassandra
 - ✦ Secrets - Hashicorp Vault



About those APIs

S3 Compatible API

- ↳ Implements the 38 most commonly used operations from the 97 in the Amazon S3 API
- ↳ Works with the AWS CLI, all of the S3 SDKs, most off-the-shelf S3 client apps
 - ✦ Just use Backblaze application key & id and set 'endpoint URL' to the value for your bucket, e.g. `https://s3.us-west-001.backblazeb2.com`
- ↳ Client must sign *every* API call

```
curl -H 'Authorization: AWS4-HMAC-SHA256 \
Credential=AKIAIOSFODNN7EXAMPLE/20130524/us-east-1/s3/aws4_request, \
SignedHeaders=host;x-amz-date, \
Signature=f0e8bdb87c964420e857bd35b5d6ed310bd44f0170aba48dd91039c6036bdb41' \
-H 'x-amz-date: 20130524T000000Z' \
https://s3.amazonaws.com/
```

B2 Native API

- Simple RESTful API

- Client exchanges credentials for an authorization token and API URL

```
curl -u "<application key id>:<application key>" \  
    https://api.backblazeb2.com/b2api/v2/b2_authorize_account  
{... "authorizationToken": "<auth token value>",  
    "apiUrl": "https://api123.backblazeb2.com", ...}
```

- Client includes authorization token value in the Authorization HTTP header in subsequent calls to the API URL

```
curl \  
    -H "Authorization: <auth token value>" \  
    -d '{"accountId": "<account id>"}' \  
    https://api123.backblazeb2.com/b2api/v2/b2_list_buckets
```

B2 Native API - Uploading Files

🔗 Two-step process:

1. Client requests an upload URL

```
curl -H 'Authorization: <auth token value>' \  
  -d '{"bucketId": "<bucket id>"}' \  
  https://api.backblazeb2.com/b2api/v2/b2_get_upload_url
```

```
{... "uploadUrl" : "https://pod-000-1005-  
03.backblaze.com/b2api/v2/b2_upload_file?cvt=c001_v0001005_t0027  
&bucket=4a48fe8875c6214145260818",  
  "authorizationToken" : "<upload auth token value>", ...}
```

B2 Native API - Uploading Files

- Client POSTs data to that URL

```
curl -H "Authorization: <upload auth token value>" \  
  -H "X-Bz-File-Name: <file name>" \  
  -H "Content-Type: <content type>" \  
  -H "X-Bz-Content-Sha1: <sha1 of file>" \  
  --data-binary "@<path to local file>" \  
  https://pod-000-1005-03.backblaze.com/b2api/b2_upload_file...
```

Every pod is accessible from the Internet!

Scale: Step 4 - Geographic Regions

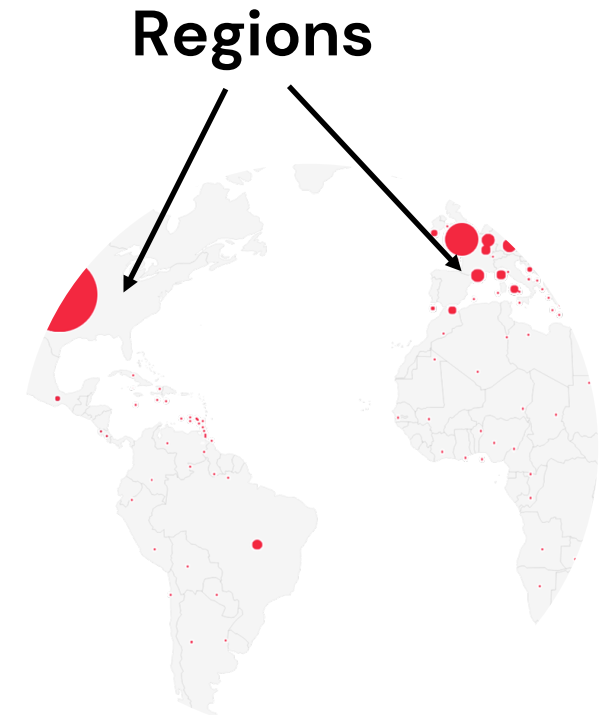
The only layer exposed to customers is the geographic region.

Geographic Region - Multiple Clusters

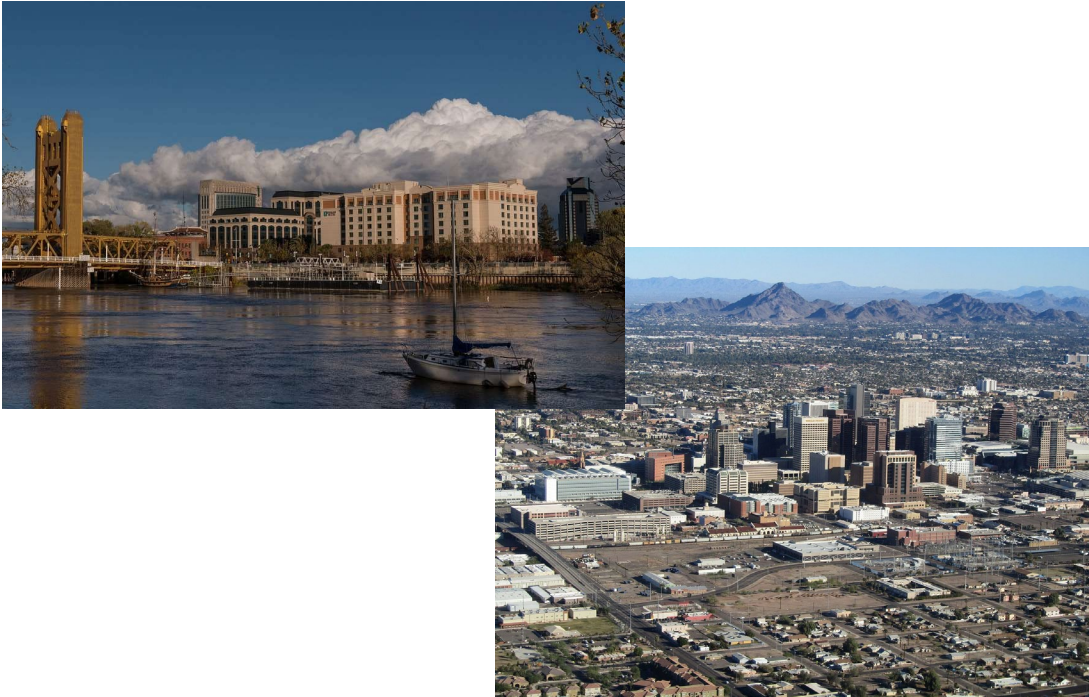
Backblaze Cluster 100 vaults
 2,000 pods
 120,000 hard drives

Backblaze Vault - 20 pods

Backblaze Storage Pod - 1 server, 60 drives



Currently Two Regions



US West: 4 Clusters

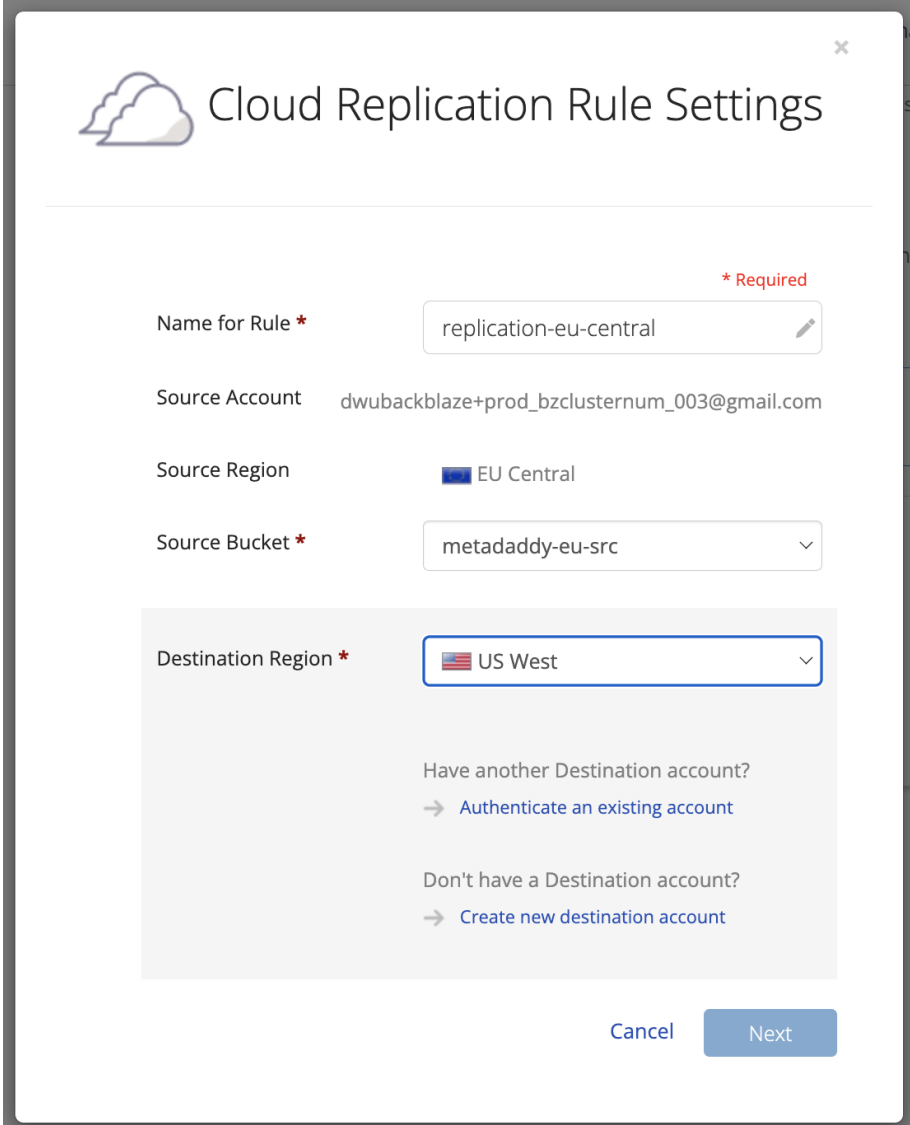
- 📍 Sacramento, CA
- 📍 Phoenix, AZ

Europe: 1 Cluster

- 📍 Amsterdam, Netherlands

Cloud Replication

- 🔗 Use cases:
 - ✦ Protecting data for security, compliance, and continuity purposes.
 - ✦ Bringing data closer to distant teams or customers for faster access.
 - ✦ Providing version protection for testing and staging in deployment environments.
- 🔗 Configure replication rules via UI or API

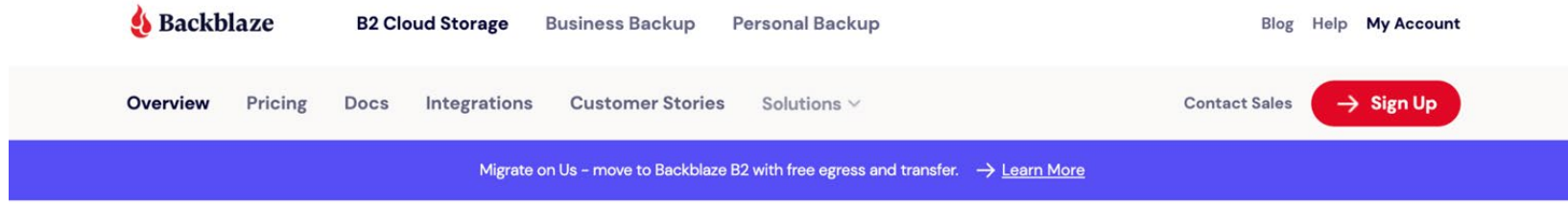


The screenshot shows a 'Cloud Replication Rule Settings' dialog box with the following fields and options:

- Name for Rule *** (Required): A text input field containing 'replication-eu-central'.
- Source Account**: A text input field containing 'dwubackblaze+prod_bzclusternum_003@gmail.com'.
- Source Region**: A dropdown menu with 'EU Central' selected.
- Source Bucket *** (Required): A dropdown menu with 'metadaddy-eu-src' selected.
- Destination Region *** (Required): A dropdown menu with 'US West' selected.
- Below the Destination Region dropdown, there are two options:
 - 'Have another Destination account?' with a link '→ Authenticate an existing account'.
 - 'Don't have a Destination account?' with a link '→ Create new destination account'.
- At the bottom right, there are 'Cancel' and 'Next' buttons.

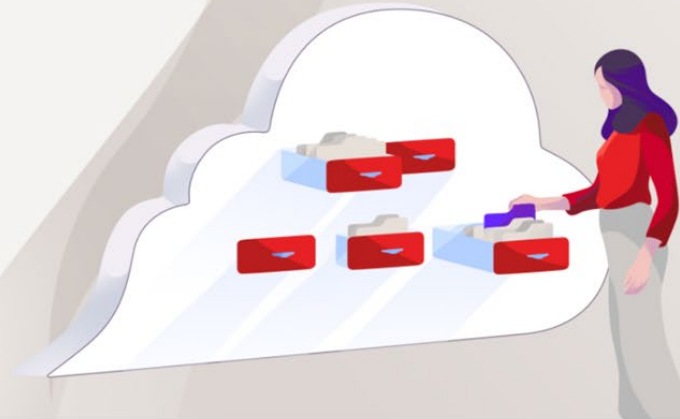
Get Started Today with 10 GB Free of Charge

No credit card necessary; no salesperson will call



Simple, Scalable Cloud Storage Services

Grow on object storage made for building applications, safeguarding data, and doing more with ease.



backblaze.com/b2

Questions?

STORAGE DEVELOPER CONFERENCE



Fremont, CA
September 12-15, 2022

BY Developers FOR Developers

A **SNIA** Event

Thank You!

Pat Patterson

Chief Technical Evangelist, Backblaze

@metadaddy / pat@backblaze.com



Please take a moment to rate this session.

Your feedback is important to us.