ANALYSIS · STORAGE · ECONOMICS

Data provenance at world scale: every object proven on chain, its bytes kept off it

Every object a store holds can carry a public, permanent proof of what it is, who wrote it, when it existed, whether it is still held, and whether a listing that names it is complete, all checked against block headers alone. bstore gets there by splitting storage into two flows: the bytes ride the data carrier and are never mined, while a proof of a few hundred bytes to a few kilobytes is. This paper measures that split from 1 KiB to 4 GiB, prices it, and scales it to the world: every new object Amazon S3 stores, 80 trillion a year, could be proven on chain for about $1.85 billion a year at today's miner fee, about two thousandths of a cent each. It closes with how far that reaches, from one object to the whole storage market.

Author: Jeff Harris · Lightweb Inc. · jeff@lightweb.net

Paper 016 · **Version** 03

Audience: storage engineers, Teranode operators and miners, and the business leaders who buy storage, sell it, or depend on knowing where data came from.

We = Lightweb Inc., designer of bstore and operator of the 1BSV network; **you** = a buyer, a provider or a miner. Lab figures are from one closed-loop run on a private chain; every projection is scenario arithmetic on cited inputs, not a forecast.

80 T	$0.000023	14 KB	$1.85 B
objects a year	**per object**	**on chain for 1 GiB**	**a year at today's fee**
every new S3 object, proven	the whole proof, paid once	the bytes never touch a block	$18.5 M at 1 sat per kB

Figure 1 - the headline: a proof for every object the largest object store adds, at scenario scale and today's fee.

The short version

- **Provenance for every object, not just the valuable ones.** A put mines a proof that ties every 4 KiB segment of an object to its writer's key, a block and its bucket's listing. Anyone can check it, at any later time, with block headers and nothing else.
- **The world's new objects fit.** S3 adds about 80 trillion objects a year. Proving every one would cost about $1.85 billion a year at today's fee of 100 sat per 1,000 bytes, 1 to 2.5 percent of what the world spends on cloud storage, and about $18.5 million at 1 sat.
- **A proof costs what S3 charges to keep one megabyte for one month.** Weighted for a real size mix, an object's whole proof is about 1,430 bytes of block: 143 sat, $0.000023, paid once.
- **The chain carries proofs, not bytes.** A run that stored 24.0 GB in 24,825 objects added 26.7 MB of blocks. A 1 GiB object mines about 14 KB, and the ratio levels off near 80,000 to 1.
- **The bytes stay retractable; the record does not.** A writer can sweep an object's bytes off every host, while the proof that it existed, and when, stays on chain.
- **Chain load follows objects.** If 1 percent of S3's new objects were stored this way, the chain would carry about 14,500 transactions a second and 3.1 TB of blocks a day: 1.4 percent of the rate Teranode sustained in its public test.
- **It changes every level of the market.** Objects become evidence, listings become proofs, providers are paid only on proof, buyers switch without moving a byte, and miners earn on the world's data without storing it.
- **Familiar today, native tomorrow.** Ordinary S3 tools write through a gateway today. A native writer removes the gateway's copies; the proof and its price do not change.

1. What one object's proof shows

A provenance claim is only as good as what a stranger can check without asking anyone. Most systems answer "where did this come from?" with a signature from a party you must trust, a registry someone else runs, or a log its owner could rewrite. bstore answers it with a chain of commitments that ends in a block header.

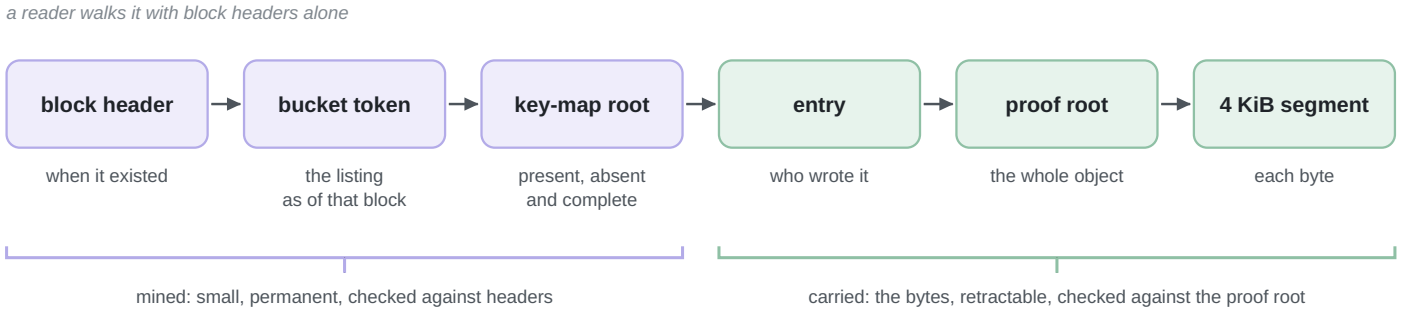

Figure 2 - the chain of commitments a reader walks for one object, from a block header to any byte, and what each link proves.

Question	What answers it	Checked against
What is it?	the proof root: one fingerprint that commits to every 4 KiB segment	the bytes themselves
Who wrote it?	each carrier spends a mined funding output that only the writer's key can spend	the writer's public key
When did it exist?	the bucket token that commits to it is mined in a block: the object existed no later than that block	the block header
Is the listing honest?	each token commits to the root of the bucket's whole key map, in the tree construction Certificate Transparency uses	block headers alone: a key's presence, its absence, a page's completeness
Is it still held?	a paid challenge samples segments chosen from a recent block, so a provider can answer only by holding the bytes	the proof root; the outcome can be published on the bucket
Can it be erased?	the bytes, yes: sweeping the funding outputs removes the carriers from every host	the commitment stays: the record that it existed, and when, cannot be removed

Every object gets this, at the same price, whatever it is worth. That makes provenance a property of the store rather than a product bought for a few important files. The training set behind a model, the original of a photograph, the evidence in a dispute, the dataset behind a published result, an audit log: each can be shown complete, unaltered and dated, by anyone, at any later time.

2. Two flows

Keeping data on a blockchain has a reputation for scaling badly, because it usually means mining the data: every byte pays a miner, and the bigger the objects, the fewer fit in a block. bstore separates two flows.

- **The bytes ride the data carrier.** Each 4 MiB piece of an object is a small transaction that is never mined, published once on the network and delivered to every provider on the bucket. The fabric is numbered and loss-free; a provider verifies each piece against the object's proof root.
- **The proof is mined.** Each carrier spends one output of a mined funding tree. A put mines a hold (the writer's bounty, offered to the providers that prove they hold the object) and a settlement (the bounty paid by rank to the first to pass a challenge). The bucket's tokens commit to each entry.

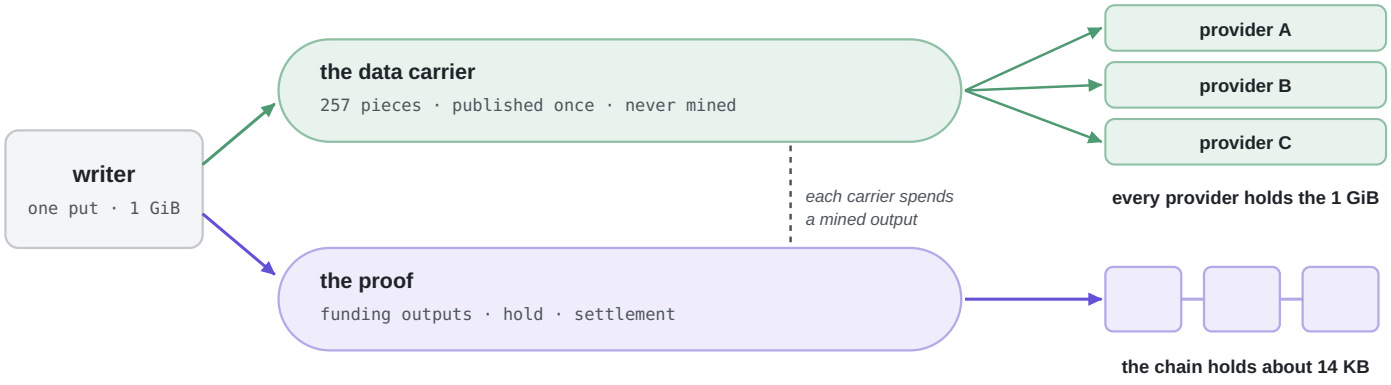

Figure 3 - the bytes and the proof travel apart; each carrier spends a mined output, so the two stay linked.

3. What the lab measured

The sweep wrote objects from 4 GiB down to 1 KiB through the S3 gateway to two providers, each object distinct and deleted after its step, in a closed loop of 1 to 16 workers. Teranode's block-assembly counter, checked against the mined blocks, gave each step's chain transactions.

Object size	Objects	Chain tx per object	Block bytes per object, measured[1]	Proof per object, structural[2]	Stored bytes per mined byte[2]
4 GiB	1	**1.00**	**50.7 KB**	**54.2 KB**	**79,231**
1 GiB	3	2.00	none in window	14.1 KB	76,347
256 MiB	12	1.67	4.7 KB	4.1 KB	65,793
64 MiB	48	1.10	2.1 KB	1,584 B	42,367
16 MiB	192	0.71	1,102 B	960 B	17,476
4 MiB	768	0.52	1,119 B	804 B	5,217
1 MiB	2,048	0.37	1,074 B	804 B	1,304
256 KiB	4,096	0.43	1,063 B	804 B	326
64 KiB	4,836	0.41	1,063 B	804 B	82
16 KiB	4,553	0.44	1,075 B	804 B	20
4 KiB	4,287	0.43	1,072 B	804 B	5.1
1 KiB	3,981	0.44	1,062 B	804 B	1.3
Run	**24,825**	**0.43**	**1,075 B**		**900**

1. Bytes of the blocks mined during the step, divided by its objects: funding trees, holds and the bucket's tokens. Settlements are not in it, because each object was deleted after its step, which retires its race; the 1 GiB step's transactions were mined outside its window. 2. One 52-byte funding output per piece and one for the entry, plus about 650 bytes for the hold and a three-payee settlement.

The chain's load is set by how many objects are written, not how many bytes. Up to 4 MiB the proof is a fixed 0.8 KB. Above it, the proof grows by one 52-byte output per 4 MiB piece, so the ratio levels off near 80,000 to 1.

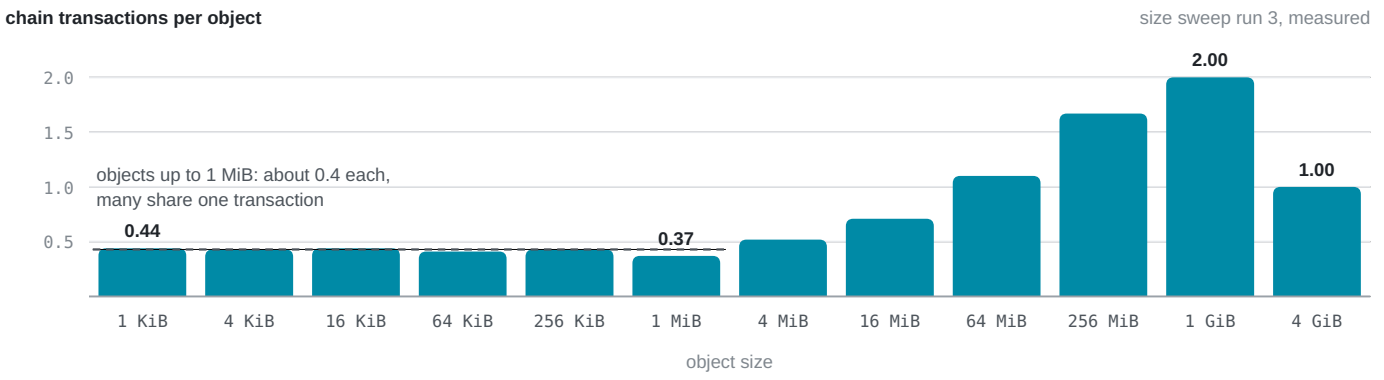

Figure 4 - chain transactions per object by size, measured: a small object shares a transaction, a 4 GiB object needed one.

Transactions per object are not bytes per object. Funding outputs come from trees of up to 1,024 outputs minted ahead of the puts, so a step's count holds its objects' holds plus any trees minted during it. The 4 GiB step's one transaction was 50.7 KB, close to the 54.2 KB its proof needs. Measured bytes run above the structural count for small objects, about 1,070 against 804, because the bucket's tokens and the trees' own overhead land there too.

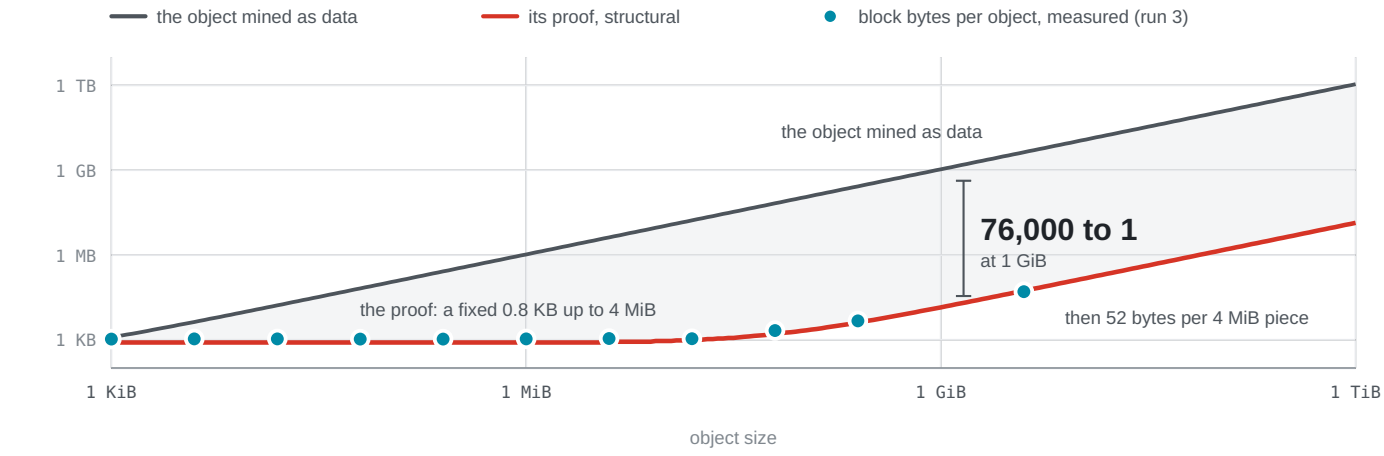

Figure 5 - bytes mined per object against its size: the object mined as data, against its proof (structural line, measured points).

4. What it costs on chain

Every number here is a measured chain cost or arithmetic at a stated provider price. bstore's software sets no price: providers publish their terms and writers choose their offers. Rate: 1 BSV = 16.14 USD = 14.52 EUR (2026-11-11, 13:29 UTC). Miner fee: 100 sat per 1,000 bytes, the published policy of two of the network's broadcasters on the same day.

The ledger is one measured put of 1 GiB to three providers for 30 days, at the provider software's default price.

Who	What	sat	USD	EUR
writer	funding outputs: 257 pieces and 1 entry, fee share	1,269	0.000205	0.000184
writer	funding outputs' value, 1 sat each	258	0.000042	0.000037
writer	hold output's transaction	100	0.000016	0.000015
writer	settlement transaction	100	0.000016	0.000015
writer	bounty, paid by rank to 3 providers	22,500	0.003632	0.003267
writer	**the writer's whole put**	**24,227**	**0.003910**	**0.003518**
others	the bucket's two tokens, three providers' advertisements	500	0.000081	0.000073
reader	one paid 1,024-sample challenge, to the provider	1,034	0.000167	0.000150
all	**fees to miners**	**1,969**	**0.000318**	**0.000286**
all	**paid to providers**	**23,534**	**0.003798**	**0.003417**

1. Measured ledger of a single put on a private chain at 100 sat/kB. Mining the same gigabyte as data at the same fee would cost 107,374,183 sat ($17.33, €15.59): 54,500 times this put's miner fees.

Proving a gigabyte costs three hundredths of a cent in fees; keeping it costs what providers ask. At the default provider price (250 sat per GiB-day, a starting point and not a market price), three copies cost about a sixth of S3 Standard's list price per GB-month.

The honest edge is the small object. Every object pays its own funding, hold and settlement, about 210 sat ($0.000034), and at least a provider's minimum bounty (1,000 sat by default), so for keeping bytes S3 Standard is cheaper by the GB below about 9 MB. For proving them, size does not matter: the proof is the same 0.8 KB at 1 KiB as at 4 MiB.

5. The world's data, proven

What would it take to prove every object the world writes? Three inputs: how many objects, how many chain bytes each, and how they split by size.

- **Objects written.** Amazon S3 stored over 100 trillion objects in 2021 and more than 500 trillion in March 2026, while serving more than 200 million requests a second. The net growth alone is about 2.5 million objects a second, 80 trillion a year: a floor on S3's writes (deletes and overwrites are not counted), and one provider's.
- **The size mix.** In IBM Cloud Object Storage traces, 85 percent of objects were under 1 MB and held 3 percent of the bytes.
- **Chain bytes per object.** Measured: 0.41 transactions per small object, about 1.5 per large one, 2,492 bytes per transaction. Weighted 85 to 15, that is 0.57 transactions and 1,429 bytes of block per object: 143 sat at today's fee.

Share of S3's new objects	Objects a year	Chain tx/s	Blocks per day	Fees per day[1]	Fees per year[1]	Of Teranode's tested rate
0.01%	8.0 B	145	31.3 GB	$505	$0.18 M	0.01%
0.1%	80 B	1,454	313 GB	$5,053	$1.8 M	0.14%
1%	**800 B**	**14,538**	**3.1 TB**	**$50,533**	**$18.5 M**	**1.4%**
10%	8.0 T	145,385	31 TB	$505,333	$185 M	14%
100%	**80 T**	**1,453,849**	**313 TB**	**$5.05 M**	**$1.85 B**	**140%**

1. At 100 sat per 1,000 bytes and 16.14 USD per BSV; provider bounties are on top and set by providers. For scale, mainnet carried 163,224 transactions (1.9 a second) and 0.41 GB of blocks in the 24 hours to 2026-11-09; Teranode sustained an average of 1.04 million transactions a second over a two-week public test.

For $18.5 million a year, 800 billion objects a year would carry a public proof of origin, integrity, time and listing. The bytes stay off chain: at S3's implied mean of 0.2 to 2 MB an object, the 1 percent row puts 0.4 to 4.4 PB a day on providers while the blocks grow by 3.1 TB.

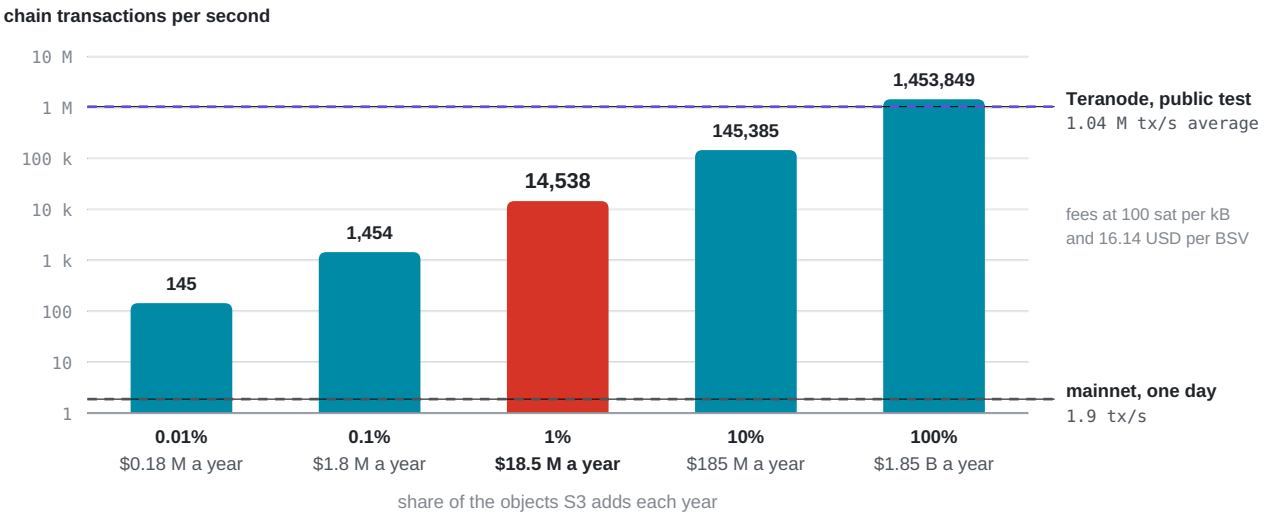

Figure 6 - chain transactions per second if a share of S3's new objects were proven on chain, against today's mainnet and Teranode's tested rate.

The fee is the lever. The table uses today's published policy, but miners set the fee in satoshis, and one peer already advertises 1 sat per 1,000 bytes.

Every new S3 object, proven	Per object	Per year	Of cloud storage spend, 2025[1]
at 100 sat/kB (today's policy)	$0.000023	$1.85 B	1.1 to 2.5%
at 10 sat/kB	$0.0000023	$185 M	0.1 to 0.25%
at 1 sat/kB	$0.00000023	$18.5 M	0.01 to 0.025%

1. Analyst estimates of $75 to $161 billion for 2025, a wide range by definition. At 100 sat/kB the 100 percent row would pay about 114 million BSV a year, more than the 21 million that will ever exist: at that scale the fee in satoshis must fall, so read the dollar column.

Provenance for every new object in the world's largest object store would cost 1 to 2.5 percent of world cloud-storage spend at today's fee, and far less as fees fall. The tables are not a forecast of share, price or date. They are the load and the fee the design would place on the chain at each share, from one run's measured ratios and one provider's public object count. The size mix moves them little: with all small objects the 1 percent row falls to 10,400 transactions a second; with half large objects it rises to 24,000.

6. Disruptive at every level

Level	Today	With every object proven on chain
An object	a file on someone's disk; its origin is whatever the metadata says	evidence: who wrote it, when, every byte, and whether it is still held, checked by anyone from block headers
A listing	a provider's report of what a bucket holds	a proof of a key's presence, its absence, and a page's completeness
A buyer	pays by the GB-month on one company's word; leaving costs egress	buys copies from competing providers and checks them; switches by changing an offer, not by moving a byte
A provider	sells capacity under a brand, trusted by contract	anyone with a disk earns per object, paid only on proof; its record of passed challenges is public
A miner	fees follow payments; data on chain means mining its bytes	fees on the world's data without storing it: one proof per object, never per byte
A data market	provenance bolted on, silo by silo, through registries, certificates and audits	one public ledger links every object; data is traded and audited by proof
The world	proving where data came from is a premium service for a few files	a property of storage itself, for about $0.000023 an object at today's fee

Provenance stops being a niche. Analysts sized data marketplace platforms at $1.9 billion and data monetization at $3.9 billion in 2025; both rest on buyers trusting where data came from. A proof that costs what S3 charges to keep a megabyte for a month makes that trust a default, not a purchase.

7. What the lab moved

This section carries rates and can be read apart from the rest: none of the ratios or costs above depend on it.

Two users' puts are delivered to every provider at 1.7 gigabytes a second, held for over a minute, and no datagram is lost on the way. Delivered means both providers hold every piece and every record of every object.

Load	Delivered to every provider[1]	Held for	What it counts
One user	**1,137 MB/s**	66 s	280 objects of 256 MiB, 70 GiB, every put answered, every piece on both providers
Two users at once	**1,721 MB/s**	68 s	1,075 MB/s and 646 MB/s, 108 GiB in 432 objects, each user with its own bucket, storage node and providers; every piece on both providers of each, and none of the fabric's 13 million datagrams lost
Uploads answered by one gateway	1,483 MB/s	6 s	32 S3 PUTs of 256 MiB, 16 at once

1. Unique 256 MiB objects through the S3 gateway, 16 at once (8 for the second user), on 10 Gb/s links, each deleted behind the load so that a run can outlast the stores. The stores are held in memory, so the figures are the path's and not a drive's. The one-user run was repeated at 1,090 and 1,142 MB/s; the two-user run came two days later and was repeated at 1,714 and 1,697 MB/s. That night one user alone delivered 1,063 to 1,070 MB/s, its storage node setting the pace.

A user's stream is one flow, and the fabric grows by flows. A bucket is one topic, and a topic travels as one flow from its storage node's lane to every provider: at most 1,246 MB/s on the wire of one 10 Gb/s link. A second user brings a bucket, a lane and flows of its own. A user that outgrows one link publishes over several lanes, each a numbered flow: two lanes carried one user's 1,067 MB/s across two links, at three fifths of each.

What sets the pace now:

- **The storage nodes.** A put is hashed six ways (blobs, content, segments, field signature, outputs, transaction) and spooled before it is minted; a native writer removes the spool (section 8). The first user's node set its pace: paced faster, a node refuses uploads.
- **The hosts behind the second user.** The second user ran on smaller hosts at about three fifths of the first's rate; pushed harder, its receiving host falls behind before the fabric does.
- **Not the fabric.** No datagram was lost in the two-user runs, and the hosts the users share, the edge that fans the pieces out and the listener that delivers them, were not what refused or fell behind.

By object size, from the size sweep rerun on the build of the one-user run (a closed loop of 1 to 16 workers, about 3 GiB a step):

Object size	Delivered to both providers[1]	Earlier build
4 GiB	566 MB/s	341 MB/s
1 GiB	645 MB/s	408 MB/s
256 MiB	680 MB/s	427 MB/s
64 MiB	774 MB/s	438 MB/s
16 MiB	565 MB/s	441 MB/s
4 MiB	230 MB/s	267 MB/s
1 KiB to 1 MiB	59 to 69 objects/s	66 to 83 objects/s

1. One gateway, one storage node and one provider host. Each step is a few seconds long, so it includes its own start; the sustained figures are the table above. The small-object rows are the loop's own pace, not a ceiling. The ratios of section 3 are the earlier run's.

None of these is the chain. In every step of both sweeps the chain carried 0 to 35 transactions a second while the storage node was saturated, and a gigabyte a second of 256 MiB objects is about four objects a second, each with its proof.

8. Familiar today, native tomorrow

The S3 gateway lets today's engineers, tools and applications use bstore without changing a line: a bucket, a key, a PUT. That familiarity has a cost, measured on the storage node, and none of it touches the proof.

Gateway cost	What a native writer avoids
Two phases in series: the whole upload is spooled, then pieces are minted and sent	the writer mints as it reads its own bytes, and pieces leave as they are made
A second home for every byte: the gateway alone touches DRAM about 31 times per delivered byte	no spool and no second copy; the node only stores
One TLS flow per upload, one identity's holds per gateway	many writers, each with its own identity and funding

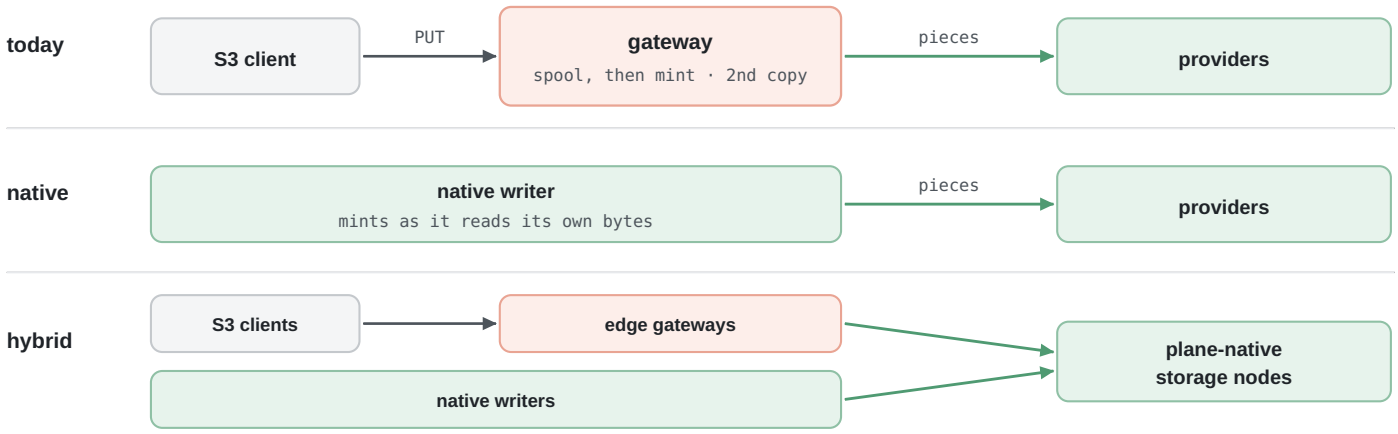

Figure 7 - the same flows and the same proof, with the gateway moved from the middle of the path to its edge.

- **A native writer.** The bstore client as a library and a command, writing pieces onto the network itself; capacity grows with writers.
- **A plane-native storage node.** Pieces from the network straight to disk in one pass, so memory and cores go to storing rather than copying.
- **A hybrid.** S3 gateways at the edge, native writers inside, the same nodes behind both. Each user picks familiarity or speed; the proof is the same.

Further out, subscription by topic will let a provider take only its share of a bucket's pieces. Topics are addressed by SHA-256 and a subscriber takes the ranges it asks for, so each provider's traffic stops growing with the whole store's write rate, and capacity grows by adding providers.

9. Assumptions and limits

Quantity	Value	Basis
Chain tx per object, by size	0.37 to 2.00	**measured**: size sweep run 3, closed loop, private chain
Block bytes, run	26.7 MB for 24.0 GB; 1,075 B per object	**measured**: same run, 10,712 transactions
Proof per object	52 B per output + about 650 B	**structural**: funding output, hold and settlement sizes
Provenance properties	proof root, key, token, key map, challenge	**structural**: as Pattern 006 describes them
1 GiB put ledger	24,227 sat; 1,969 sat in fees	**measured**: one put, three providers, 30 days, default terms
Miner fee	100 sat/kB; 10 and 1 as scenarios	**verified**: two broadcasters' policy, 2026-11-11; one peer advertises 1 sat/kB
BSV rate	16.14 USD, 14.52 EUR	**verified**: CoinGecko, 2026-11-11 13:29 UTC
S3 objects	100 T (2021), 500 T (2026)	**verified**: AWS News Blog, 2021 and 2026-03-13
Small-object share	85% under 1 MB	**stated**: IBM COS traces (2019); no industry-wide median is published
Cloud storage spend	$75 to $161 B (2025)	**modeled**: analyst range
Mainnet load	163,224 tx in 24 h	**verified**: WhatsOnChain, one day to 2026-11-09
Teranode tested rate	1.04 M tx/s average	**stated**: BSV Association public test, 2025
Rates	one user 1,137 MB/s for 66 s; two users 1,721 MB/s for 68 s; by size 230 to 774 MB/s	**measured**: sustained runs two days apart and one size sweep, memory-backed stores

Limits, stated plainly:

1. **One run on a private chain.** The ratios come from one closed-loop run with two providers. They match the structural count, but a public network with many writers has not been measured.
2. **S3's net growth is a floor and one provider.** Writes include overwrites and deletes that net growth hides, and other clouds publish no object counts. Read the scenario rows as orders of magnitude.
3. **The 100 percent row is beyond tested capacity.** It asks 140 percent of the rate Teranode has shown, and more satoshis a year than will ever exist at today's fee. It sizes the problem; it promises nothing.
4. **No market exists yet.** We run no public provider. Every provider price here is the software's default, not a quote.
5. **Provenance proves the record, not the truth of the content.** A proof shows who wrote which bytes and when; it cannot show that what they wrote is accurate.
6. **Small objects cost more to keep than at S3.** Bundling is the answer; the scenarios count objects as written, not bundled.

References

bstore: storage as a market, where providers race for your data and are paid only on proof (Lightweb Inc., Pattern 006): https://1bsv.net/patterns/bstore.pdf

The data carrier: proven without being mined, and unmineable because everyone receives it (Lightweb Inc., Paper 011): https://1bsv.net/papers/data-carrier.pdf

Sharding the object plane (Lightweb Inc., Paper 010): https://1bsv.net/papers/object-plane-sharding.pdf

The overlay object plane (Lightweb Inc., Paper 006): https://1bsv.net/papers/overlay-object-plane.pdf

Certificate Transparency (IETF, RFC 6962): https://www.rfc-editor.org/rfc/rfc6962

Twenty years of Amazon S3 and building what's next (AWS News Blog, 2026-03-13): https://aws.amazon.com/blogs/aws/twenty-years-of-amazon-s3-and-building-whats-next/

Amazon S3's 15th birthday (AWS News Blog, 2021): https://aws.amazon.com/blogs/aws/amazon-s3s-15th-birthday-it-is-still-day-1-after-5475-days-100-trillion-objects/

It's Time to Revisit LRU vs. FIFO (Eytan et al., HotStorage 2020): https://www.usenix.org/system/files/hotstorage20_paper_eytan.pdf

Amazon S3 pricing (AWS): https://aws.amazon.com/s3/pricing/

ARC policy (GorillaPool): https://arc.gorillapool.io/v1/policy

ARC policy (TAAL): https://arc.taal.com/v1/policy

Mainnet fee floor (bsv-blockchain/broadcaster, issue 398): https://github.com/bsv-blockchain/broadcaster/issues/398

Chain and block API (WhatsOnChain): https://api.whatsonchain.com/v1/bsv/main/chain/info

Public release of Teranode (BSV Association, 2025-10-14): https://www.prnewswire.com/news-releases/bsv-association-bsva-announces-public-release-of-teranode-software-302583362.html

BSV Blockchain network topology and Teranode dynamics (BSV Blockchain Hub): https://hub.bsvblockchain.org/higher-learning/bsv-academy/deep-dive-in-bsv-blockchain/bsv-blockchain-network-topology-and-teranode-dynamics/bsv-blockchains-network-topology

BSV rate (CoinGecko API): https://api.coingecko.com/api/v3/simple/price?ids=bitcoin-cash-sv

Cloud storage market (Fortune Business Insights): https://www.fortunebusinessinsights.com/cloud-storage-market-102773

Data marketplace platform market (Grand View Research): https://www.grandviewresearch.com/industry-analysis/data-marketplace-market-report

Data monetization market (Grand View Research): https://www.grandviewresearch.com/industry-analysis/data-monetization-market

---

1BSV - BSV Layered Multicast - cash moves fast on 1BSV. · 1bsv.net

---

Lightweb Inc. · Jeff Harris · jeff@lightweb.net · github.com/lightwebinc

Unbounded Solutions for a Small World™ · © Lightweb Inc.

© 2026 Lightweb Inc. All rights reserved. 1BSV and Unbounded Solutions for a Small World are trademarks of Lightweb Inc. This paper describes an architecture in rapid development: the design, its interfaces and this document are subject to change without notice, and later versions supersede this one. Projections are scenario arithmetic, not forecasts. Nothing in this paper is a commitment to ship any feature, an offer of service, or professional advice.