BLOCK STORAGE

maxRAID Grid

WildFire Storage

Block storage for cloud and AI providers

Fast. Efficient. Keeps running when servers fail.



A 10-server grid with one spare server: each 8+2 array (eight data drives, two parity) puts one drive in every server, on its own diagonal. If a server fails, the spare takes over its array.

Lower SSD cost

About 60% less flash than keeping three copies. Low SSD wear. Compression, and space used only when written.

20 GB/s per logical array

And 2 million operations per second, read and write. No central bottleneck. Up to 200 GB/s on a 10 × 10 grid, with 200 GbE to every server.

Survives two failures

Each 8+2 array survives any two failed drives, or two whole servers, and repairs itself automatically.

Cloud and AI ready

Many customers, per-customer limits (QoS), NVMe-oF over TCP or RDMA, GPUDirect and iSCSI.

Grows to petabytes

Use very large QLC SSDs to reach multi-petabyte capacities.

Your hardware

Standard x86 servers; SATA, SAS or NVMe SSDs (TLC, QLC, PLC). Modest memory needs.

A grid of arrays

maxRAID Grid is a grid of logical arrays. Each one is a parity-protected array (K data and M parity drives, erasure-coded) with its own software flash layer, but its drives are spread across the grid: one drive in every server, on its own diagonal. Clients connect to volumes, and each array's volumes are served by one server, its master. A 10 × 10 grid has ten arrays, and each server is the master of one, so client traffic is spread across every server. The servers share a fast storage network (multi-path NVMe-oF), so losing a drive, or a whole server, costs each array at most one drive. The protection comes from the grid itself, without buying a second copy of every drive.

Native performance. Each array runs independently, with no central metadata server, so it performs like a local array: 20 GB/s and 2 million operations per second (IOPS), read and write, per array. Spread volumes across arrays for as much bandwidth as your network can carry; ten servers with ten SSDs each reach 200 GB/s and 20 million IOPS, with 200 GbE to every server.

Less flash per usable terabyte. An 8+2 array leaves 80% of its flash usable; keeping three full copies leaves 33% — with the same protection against two failures, and higher performance. Inline compression, space used only when written (thin provisioning), very low flash wear and support for the newest high-capacity SSDs (QLC and PLC, including large indirection units) stretch every byte further. All of this runs on standard x86 servers with modest memory needs.

Self-healing. Failed SSDs are replaced automatically. If a server fails, another server takes over its array under the same network address, so clients carry on without reconfiguration. That's the spare server when there is one; otherwise a surviving server serves two arrays until the failed one is replaced.

AT A GLANCE

Performance
20 GB/s and 2 million operations per second (IOPS) per array

Protection
Parity you choose: 1, 2, 3 or more drives per array; an 8+2 array survives any two failed drives or servers

Scale
50 TB to 10+ PB

Access
NVMe-oF (TCP, RDMA, GPUDirect) and iSCSI; volumes many servers can share, for AI

For cloud providers
Many customers per system; thousands of volumes per server; snapshots, backups; per-customer limits (QoS), burstable

Hardware
Standard x86 servers and Ethernet; SATA, SAS or NVMe SSDs

CLOUD OBJECT STORAGE (S3)

maxObject

WildFire Storage

Object storage built for hard drives

Built for cloud providers. Compared with Ceph RGW, the widely used open-source object store, on the same drives:

SPACE

36%

more space with conventional (CMR) hard drives

67%

more space with high-density (host-managed SMR) drives

The densest drives

High-density SMR and conventional drives of any size, mixed in one cluster.

SPEED

1 vs 12

disk reads to fetch a small object, vs Ceph RGW

Packed objects, streamed writes

Many objects share each protected stripe, so a read usually touches one disk. Writes stream to every drive, and tiny objects never touch a hard drive.

REPAIR

<30 min

to rebuild a failed 26 TB drive, while customers keep working

Every drive helps

In a 1,000-drive cluster, 999 drives share the rebuild. Only live data is rebuilt.

No one has to act

Failed drives are rebuilt, retired and removed automatically.

Object storage built around the hard drive

Hard drives are a necessity for object storage. Combined with erasure codes — protection data spread across many drives — they give you a lot of space for the money. But how the erasure codes are implemented matters.

One protection set per object costs speed and space. Many platforms, including Ceph RGW, create a separate erasure-code set for each object. Reads suffer: to read an object, you must read every piece of its set. Space is wasted because the sets are narrow and vary in size. maxObject uses erasure codes differently.

Pack objects, stream writes. maxObject packs many objects into each protected stripe, so an S3 read (GET) usually touches one disk, where Ceph RGW at 12+4 needs 12. Writes (PUTs) are collected into large regions, protected across servers in memory, and streamed to every drive. Tiny objects never touch a hard drive.

More sellable terabytes. On the same 1,000 drive slots, with the same protection against any four losses, Ceph RGW 12+4 yields 13.7 PB usable from 26 TB drives. maxObject, with wide 36+4L+3G erasure-code sets (LRC), yields 18.5 PB (36% more), and 22.8 PB (67% more) on 32 TB high-density (host-managed SMR) drives. You can mix conventional and SMR drives freely in one cluster. Ceph does not support host-managed SMR drives.

Fast repair, no downtime. The real risk in a large cluster is the time between a drive failure and full protection. maxObject keeps that window short: every healthy drive takes a share of the rebuild, so a failed 26 TB drive is rebuilt in under 30 minutes, where filling a single replacement drive would take more than a day, and the larger the cluster, the faster it finishes. Customers keep reading and writing throughout; missing data is reconstructed on the fly. If a whole server is lost, its index (metadata) is fully protected again within minutes.

Less to buy, less to run. maxObject requires under 20% of the flash and under 50% of the memory Ceph needs, and spare RAM and flash become a read cache. One program per server, self-healing, with no tuning of where data goes: simple to run and to expand.

BUILT FOR CLOUD HOSTERS

Customers
Separate accounts, with AWS-style (IAM) users and permissions

Billing
Quotas, and billing rules you define

Tiers
Flash and hard-drive tiers in one cluster, each sold by contract

Clients
AWS SDKs and CLI work as-is

Growth
Add servers or drives of any size, without a long data reshuffle

Features and figures may change. Figures are design arithmetic from published drive characteristics, not benchmarks; rebuild time assumes a 50-server cluster. Ceph is a trademark of Red Hat, Inc.

wildfire-storage.com/maxobject