



Cloud Protect365

Tech Brief

High Data Durability & Protection Capabilities

Wasabi is the hot cloud storage company delivering disruptive storage technology that is used by Cloud Protect365 to backup data with no fees for ingress, egress or API requests. You can use Wasabi hot cloud storage for a variety of purposes including on-premises or cloud-based workloads, for backup and recovery, or archival storage for long-term data retention or on-prem storage offload.

In addition to the low price and high performance capabilities, Wasabi provides a highly durable and reliable storage infrastructure, engineered to preserve data integrity and ensure high service availability. In the storage industry, durability refers to the ability of a storage platform to provide long-term protection against disk drive failures, bit rot, degradation or other corruption. Durability is generally expressed as an annual percentage rate, approaching 100%. The closer to 100%, the greater the durability of the storage platform, and the less likely you are to lose data due to drive failures, bit rot or media corruption.

Historically, premises and first-generation cloud storage vendors have achieved high durability by replicating data across multiple drives using various RAID (Redundant Array of Independent Disks) schemes. With modern public cloud object storage services such as Wasabi, the use of erasure coding is popular as a means of achieving high durability while efficiently using storage system resources.

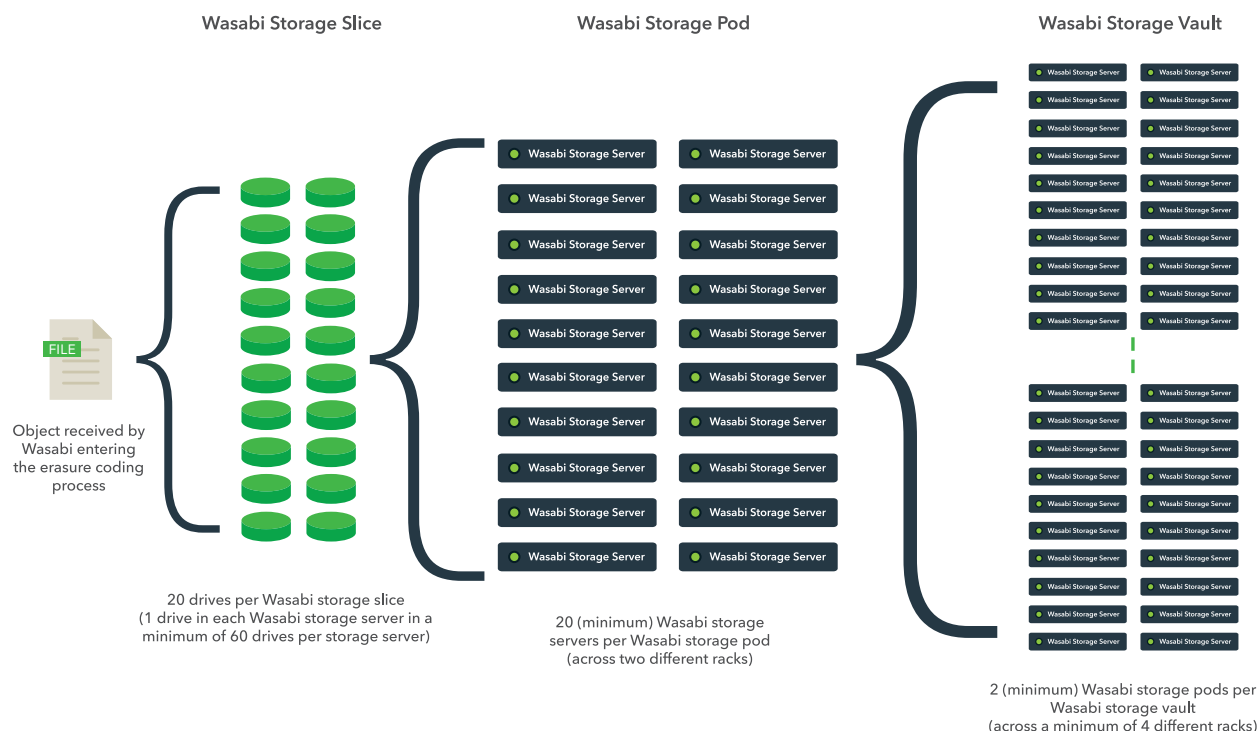
Leveraging Erasure Coding For High Data Durability

Wasabi uses advanced, industry-proven erasure coding algorithms to protect data against drive failures and associated media errors. When the service receives a customer data file (referred to as an object), Wasabi uses erasure coding to transform each object into a series of codes, which are distributed across independent drives in separate storage servers for maximum resiliency. In the event of drive failure, storage server failure or data corruption, the original data object can be reconstructed by Wasabi using only a subset of the codes.

Erasure coding is more efficient than and just as reliable as traditional replication-based data protection schemes such as RAID. Erasure coding allows Wasabi to provide 11 x 9s of data durability, fully protecting customer data without the overhead of RAID-based approaches. By leveraging erasure coding, we are able to make far more effective use of storage capacity. We pass the cost savings of our efficiencies along to our customers in the form of low pricing.

Wasabi Distributes Data Across 20 Different Drives for High Resiliency

To better understand how Wasabi uses erasure coding to achieve 11 x 9s of data durability, let's first examine the storage building blocks in Wasabi's system architecture. The most basic element in any storage platform is a disk drive. Wasabi uses 20 drives to form an element known as a Wasabi storage slice. Each of the 20 drives in a storage slice is housed in a dedicated physical server to maximize data integrity. This means a server failure impacts only a single drive (not multiple drives). In addition, the drives are distributed across two different server racks to protect against cascading power surges or other events that impact an entire server rack. Wasabi storage slices are organized into a Wasabi storage pod and Wasabi storage pods are organized into Wasabi storage vaults. Each Wasabi storage region (i.e. us-east-1, us-west-1, etc.) contains a storage vault.



Wasabi Erasure Coding Process & Storage System Elements

When an object is written to Wasabi, the Wasabi erasure coding algorithm converts the object into a series of data and parity fragments and stores each fragment on 20 different drives. When you read an object from storage, Wasabi reassembles the object using the fragments. In the event one or more drives fail (up to 4 drives), an object can be fully reconstructed using any 16 of the 20 fragments. This means that Wasabi can withstand the failure of up to any 4 drives within a storage slice without losing data.



Wasabi's use of erasure coding provides high data durability while making optimal use of storage capacity. Erasure coding also makes it easy for Wasabi to perform in-service maintenance and upgrades as drives can be swapped out without disrupting service.

Calculating Eleven 9s of Data Durability

Wasabi calculates data durability by factoring in the following metrics:

- Annualized drive failure rates (AFR)
- Mean-time-to-repair a failed drive (MTTR)
- Probability of 5 drives being unavailable at one time

For AFR, Wasabi uses industry-accepted guidelines and assumes a conservative drive AFR of 5%. In practice, our observed AFRs are much lower. For MTTR, Wasabi estimates 3.4 days per drive, based on the calculation below:

MTTR = 14 TB drive capacity * 50 MB/s drive write speed = 3.4 days to fully write a replaced 14 TB drive

As stated earlier in the erasure coding discussion, in order for an object to be lost, more than 4 drives in a Wasabi storage slice would have to fail. To understand the probability of this, the first step is to understanding the probability a single drive failing using the calculation below:

Probability of a 1 drive failing = AFR (5% year) * MTTR (3.4 days) * (1/365 year/days) = 4.66×10^{-4}

The next step would be to understand the probability of four drives failing while another drive in the storage slice was rebuilding. This would be a potential data loss scenario because five drives in a storage slice would not be available (one in rebuild mode, plus four new failures). To calculate this probability, this formula applies:

Probability of 4 drives failing = Probability of 1 drive failing (4.66×10^{-4}) to the 4th power = 4.7×10^{-14}

The final step in calculating data durability is to factor in the probability of 4 drives failing using the following formula:

Data Durability = $1 - (\text{probability of 4 drives failing}) = 1 - 4.7 \times 10^{-14} = .99999999999995$

As seen in the above calculations, Wasabi storage architecture provides greater than 11 x 9s of durability. The calculated number is actually 13 x 9s but for the sake of taking a conservative approach to the calculations and to align with how most of the hyperscalers position their data durability, Wasabi uses 11 x 9s as the published data durability metric.

One might ask how data durability of 11 x 9s translates to a real-world file loss scenario. 11 x 9s of data durability corresponds to an average expected loss of .000000001 percent of objects (1 minus .9999999999) per year. If you have 10,000,000 objects stored with Wasabi, then 11 x 9s of durability translates into 1 lost file every 10,000 years.

Data Protection Beyond Data Durability

Data durability is just one dimension of data protection and data durability should not be confused with system availability. Availability refers to the ability of a service provider to ensure continuous service in the event of system failures or catastrophes. Service level agreements (SLAs) typically include service availability (uptime) commitments. Wasabi's high system availability allows us to offer strong SLAs comparable to those offered by our competition.

The Wasabi service utilizes an architecture designed for high system availability and is hosted in top-tier data center facilities that are highly secure, fully redundant and certified for SOC 2 and ISO 27001 compliance. Each individual Wasabi data center is based on a highly scalable, fully distributed architecture with redundant system components, power sources and network connections to ensure high availability. The system design eliminates single points of failure; all system elements are protected using 1:1, 1+1 or N:M redundancy.

In addition to high data durability and system availability, the Wasabi service protects your data by leveraging these 4 features:

Data Immutability: Options for using immutability include bucket locking and object locking.

Bucket immutability works well for primary data and deep archive storage where having a set-and-forget policy at the bucket level makes it easier to prevent accidental alteration or deletion. Object immutability is well suited for backup and near-line archiving requiring more granular control of object retention settings.

Active Integrity Checking: Wasabi uses checksum mechanisms to verify the integrity of all stored objects every 90 days. We detect and repair corrupt data automatically to avoid bit rot or other corruption, mitigating risk and uncertainty immediately. This integrity checking is also applied at the time of initial object upload and everytime the object is downloaded.

Bucket Replication: Wasabi provides an optional feature known as Wasabi Bucket Replication that allows you to fully replicate the contents of one Wasabi bucket to another Wasabi bucket in the same or different geo-region. This feature allows you to implement geo-redundancy by replicating the contents (for example) of your us-east-1 bucket to a bucket in our us-west-1 data center.

Data Security: Wasabi is secure by default and all data stored in Wasabi hot cloud storage is always encrypted at rest (even if the data is already encrypted by the storage application prior to sending it to Wasabi). Wasabi follows industry-best security models and security design practices, including the AWS Identity and Access Management (IAM) model and enterprise single-sign-on (SSO). Examples of Wasabi security features include:

- HTTPS is supported for the secure upload/download of data
- Buckets are only accessible to the bucket and object creators holder



- Wasabi supports user authentication to control access to data
- Access control mechanisms such as bucket policies and Access Control Lists (ACLs) can be used to selectively grant permissions to users and groups of users

Conclusion

Wasabi provides an highly durable and reliable storage infrastructure, designed from the ground up to preserve data integrity and provide high service availability. Advanced erasure coding algorithms are used to achieve 11 x 9s of data durability and provide comparable levels of protection for a low cost. Strong security features, optional data immutability and active integrity checking provide additional protection against bit rot, administrative mishaps and tampering. A redundant system design ensures high uptime, backed by a comprehensive SLA. We can help you slash storage cost and complexity, while ensuring the integrity and availability of your mission-critical data.

To trial our solution, please contact sales@backupeverything.co.uk