Understanding Erasure Coding and Its Difference with RAID

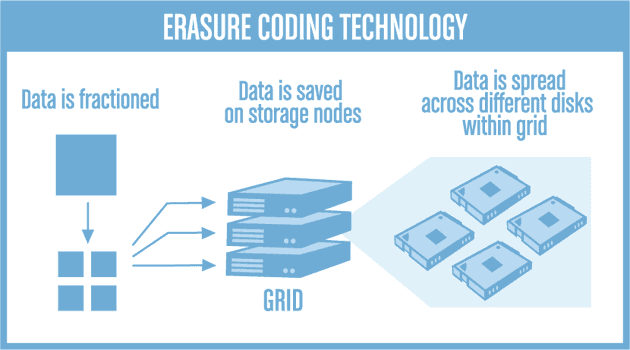
RAID is used to protect against disk failures but, many of the traditional raid levels are designed to handle only one or two disk failures in the group. With today’s larger capacity disk drives when one disk fails it can take days to rebuild data and restore processes. The ability to tolerate only one or two failures is just not enough redundancy to protect your data. Erasure coding is all about making data highly available. In simple terms it breaks data into a configurable number of parts and distributes those parts across a set of different storage systems. Let’s elaborate it:

**What is Erasure Coding?**

* [What is Erasure Coding?](https://stonefly.com/blog/understanding-erasure-coding#What_is_Erasure_Coding)
* [How Erasure Coding works?](https://stonefly.com/blog/understanding-erasure-coding#How_Erasure_Coding_works)
* [Erasure Coding VS RAID?](https://stonefly.com/blog/understanding-erasure-coding#Erasure_Coding_VS_RAID)
* [Where can Enterprises use Erasure Coding?](https://stonefly.com/blog/understanding-erasure-coding#Where_can_Enterprises_use_Erasure_Coding)
* [Benefits of Erasure Coding:](https://stonefly.com/blog/understanding-erasure-coding#Benefits_of_Erasure_Coding)

IT administrators who design storage systems must plan ahead so that mission critical data is not lost when any type of failure occurs. The storage systems comes in different shapes but they have one thing in common that is the chance of failing and losing data. The utilization of erasure coding prevents the loss of data due to system failure or disaster.

In simple words Erasure coding (EC) is one of the methods of data protection through which the data is broken into sectors. Then they are expanded and encoded with redundant data pieces and stored across different storage media. Erasure coding adds the redundancy to the system that tolerates failures.



**How Erasure Coding works?**

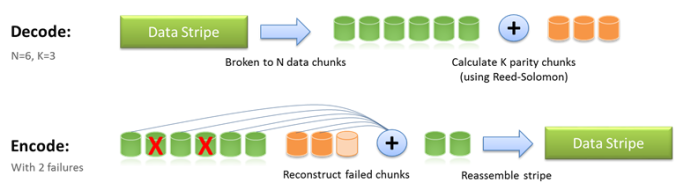
Erasure coding takes original data and encodes it in such a way that, when you require the data it only needs the subset of the pieces to recreate the original information. Let’s take an example, consider that the original value of the object or data is 95, we divide it in such a way that x= 9 and y=5. The encoding process will create a series of equations.

Suppose in this case it will create equations like:

* x+y= 14
* x-y= 4
* 2x+y= 23

When you want to recreate the object you require any two of those three equations, to be able to be decoded. So when you solve the equations together, you will be able to get the values for x and y.

We have 3 equations but we can get the original information from two of them.  The data protection scheme that breaks data into fragments, encodes them and stores them across multiple locations is Erasure coding.



**Erasure Coding VS RAID?**

Erasure coding and RAID are sometimes mixed up but they are very much different from each other. RAID allows data to be stored at different locations and it protects against drive failures. In erasure coding, the data is broken in parts, then expanded and encoded. After that the data segments are kept in multiple locations. Verily RAID facilitates data protection; however, erasure coding consumes less storage and RAID is time efficient. RAID and Erasure coding both are desirable depending on the situations.

**Where can Enterprises use Erasure Coding?**

Enterprises that require a failure free storage environment need to incorporate erasure coding technology. Here are some instances where EC can be very beneficial:

* Disk array systems
* Data grids
* Distributed storage applications
* Object stores
* Archival storage

One common current use case for erasure coding is object-based cloud storage. As erasure coding requires high CPU utilization and incurs latency, it is suited for archiving applications. Erasure coding is less suitable for primary workloads as it cannot protect against threats to data integrity.

**Benefits of Erasure Coding:**

Erasure coding provides advanced methods of data protection and [**disaster recovery**](https://stonefly.com/dr-backup-appliances). StoneFly’s appliances use erasure-coding technology to avoid data loss and bring ‘always on availability’ to organizations. Erasure coding provides numerous benefits to its users.

* **Storage Space Utilization:**Erasure coding delivers better storage utilization rate by reducing the space consumed and providing the same level of redundancy: 3 copies. Up to 50% more space is saved using erasure coding.
* **Greater Reliability:**Data pieces are fragmented into independent [fault dummies](https://stonefly.com/blog/blog/erasure-coding-innovative-method-robust-data-protection-disaster-recovery). This ensures there are no dependent or correlated failures.
* **Suitability:**Erasure coding can be used for any file size. Ranging from small block sizes of KiloBytes to large block sizes of PetaBytes.
* **Only Subsets:** You require only subsets of data to recover your data. You don’t require your original data.
* **Flexibility:** You can replace the failed components when convenient, without taking the system offline.