Erasure Coding

A Ceph pool is associated to a type to sustain the loss of an OSD (i.e. a disk since most of the time there is one OSD per disk). The default choice when [creating a pool](http://docs.ceph.com/docs/master/rados/operations/pools) is *replicated*, meaning every object is copied on multiple disks. The [Erasure Code](https://en.wikipedia.org/wiki/Erasure_code) pool type can be used instead to save space.

**We have default erasure code profile looking like this** :

**$ ceph osd erasure-code-profile get default**

**k=2**

**m=1**

**plugin=erasure**

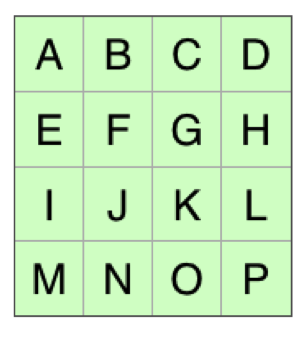
**crush-failure-domain=host**

**technique=reed\_sol\_van**

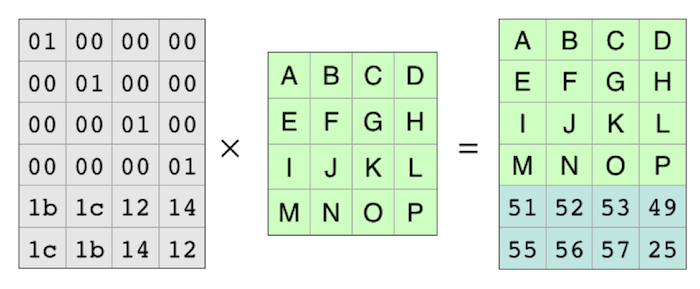
**k=2 is a data or stripping and m=1 means number of parity**

**Explaining reed\_sol\_van technique :**

In this example, the four pieces of the file are each 4 bytes long. Each piece is one row of the matrix. The first one is “ABCD”. The second one is “EFGH”. And so on.

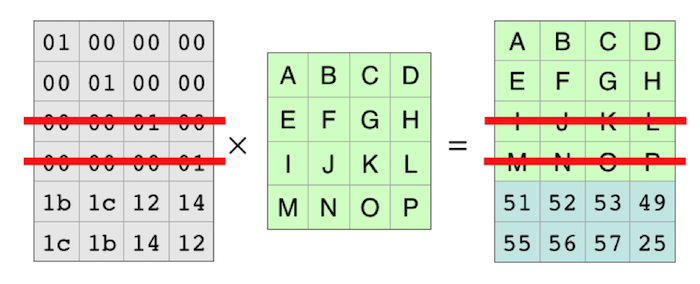


Original Data



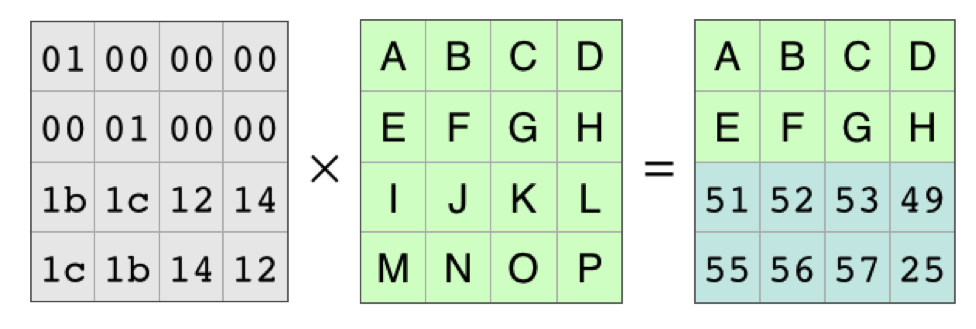
Applying the coding Matrix

The result is a matrix with two more rows than the original. Those two rows are the parity pieces.



Data:Loss 2 of the 6 rows are lost

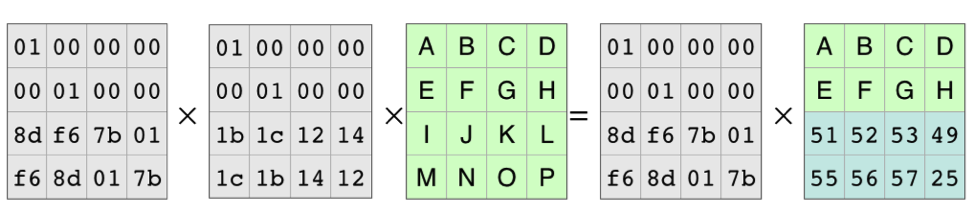
With those rows completely gone it looks like this.



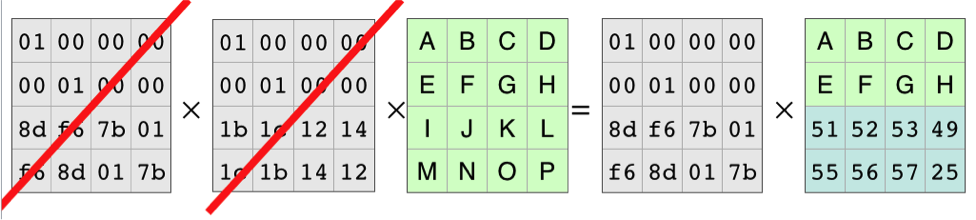
Data Loss:The Matrix without the 2 “lost” rows

Because of all the work that mathematicians have done over the years, we know the coding matrix, the matrix on the left, is invertible. There is an inverse matrix that, when multiplied by the coding matrix, produces the identity matrix.

As in basic algebra, in matrix algebra you can multiply both sides of an equation by the same thing

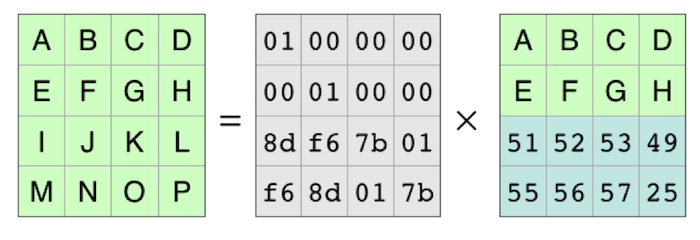


Multiplying Each Side of the Equation by the Inverse Matrix



The Inverse Matrix and the Coding Matrix Cancel Out

This leaves the equation for reconstructing the original data from the pieces that are available:



Reconstructing the Original Data

So to make a decoding matrix, the process is to take the original coding matrix, cross out the rows for the missing pieces, and then find the inverse matrix. You can then multiply the inverse matrix and the pieces that are available to reconstruct the original data.

Note: The *jerasure* plugin is the most generic and flexible plugin, it is also the default for Ceph erasure coded pools.