

EC Distribution and Recovery across Subclusters

In addition to spreading segments out across nodes (so that no two segments for the same erasure set are on the same node), Netmail Store automatically distributes segments evenly across subclusters. If it is configured correctly, this distribution keeps your data protected even if an entire site fails.

In a subcluster configuration, your data is safe when the following is true:

$$\text{Subclusters remaining} / \text{Total \# of subclusters} \geq k \text{ data-segments} / (k \text{ data-segments} + p \text{ parity-segments})$$

Example 1: Subcluster with EC Setting that does NOT Ensure Full Recovery

With two subclusters containing four nodes each and EC set to 5:2 for seven total segments, one subcluster would have three segments and the other would have four segments. The number of segments in each subcluster is greater than the parity count (in this example), so the object would NOT be fully recoverable if you lost an entire subcluster. Using the formula above to double check shows that the data is NOT safe if a single subcluster is lost:

$$1/2 \text{ IS NOT } \geq 5 / (5+2)$$

Example 2: Subcluster with EC Setting that Ensures Full Recovery

With 3 subclusters containing 5+ nodes each and EC set to 9:6 for a total of 15 segments, each subcluster would have five segments. If you lost an entire subcluster, you would only lose five segments, which is less than the set parity level, and therefore you would be able to recover the object.

Using the formula above to double check shows that the data is safe if one of the three subclusters is lost:

$$2/3 \geq 9 / (9+6)$$

The following table compares these examples.

EC Encoding	# Nodes required to write an EC object	Max # Nodes that can fail and still read the object	Protected with loss of 1 of 2 subclusters?	Protected with loss of 1 of 3 subclusters?
5:2	7	2	NO. With 7 segments, one subcluster would have 3 segments and the other would have 4 segments, both over the maximum allowed.	NO. With 7 segments, one subcluster would have 3 segments and the other two would have 2 segments each. If the subcluster with 3 objects went offline, the object would not be recoverable because that value is over the maximum allowed.
9:6	15	6	NO. With 15 segments, one subcluster would have 8 segments and the other would have 7 segments, both over the maximum allowed.	YES. With 15 segments, each subcluster would have 5 segments, which is less than the maximum allowed.