# ERASURE CODED POOL

## PURPOSE

Erasure-coded pools require less storage space compared to replicated pools. The erasure-coding support has higher computational requirements and only supports a subset of the operations allowed on an object (for instance, partial write is not supported).

## USE CASES

### COLD STORAGE

An erasure-coded pool is created to store a large number of 1GB objects (imaging, genomics, etc.) and 10% of them are read per month. New objects are added every day and the objects are not modified after being written. On average there is one write for 10,000 reads.

A replicated pool is created and set as a cache tier for the erasure coded pool. An agent demotes objects (i.e. moves them from the replicated pool to the erasure-coded pool) if they have not been accessed in a week.

The erasure-coded pool CRUSH rule targets hardware designed for cold storage with high latency and slow access time. The replicated pool CRUSH rule targets faster hardware to provide better response times.

### CHEAP MULTIDATACENTER STORAGE

Ten datacenters are connected with dedicated network links. Each datacenter contains the same amount of storage with no power-supply backup and no air-cooling system.

An erasure-coded pool is created with a CRUSH rule that will ensure no data loss if at most three datacenters fail simultaneously. The overhead is 50% with erasure code configured to split data in six (k=6) and create three coding chunks (m=3). With replication the overhead would be 400% (four replicas).

## INTERFACE

Set up an erasure-coded pool:

```
$ ceph osd pool create ecpool 12 12 erasure
```

Set up an erasure-coded pool and the associated CRUSH rule `ecrule`:

```
$ ceph osd crush rule create-erasure ecrule
$ ceph osd pool create ecpool 12 12 erasure \
    default ecrule
```

Set the CRUSH failure domain to osd (instead of host, which is the default):

```
$ ceph osd erasure-code-profile set myprofile \
    crush-failure-domain=osd
$ ceph osd erasure-code-profile get myprofile
k=2
m=1
plugin=jerasure
technique=reed_sol_van
crush-failure-domain=osd
$ ceph osd pool create ecpool 12 12 erasure myprofile
```

Control the parameters of the erasure code plugin:

```
$ ceph osd erasure-code-profile set myprofile \
    k=3 m=1
$ ceph osd erasure-code-profile get myprofile
k=3
m=1
plugin=jerasure
technique=reed_sol_van
$ ceph osd pool create ecpool 12 12 erasure \
    myprofile
```

Choose an alternate erasure code plugin:

```
$ ceph osd erasure-code-profile set myprofile \
    plugin=example technique=xor
$ ceph osd erasure-code-profile get myprofile
k=2
m=1
plugin=example
technique=xor
$ ceph osd pool create ecpool 12 12 erasure \
    myprofile
```

Display the default erasure code profile:

```
$ ceph osd erasure-code-profile ls
default
$ ceph osd erasure-code-profile get default
k=2
m=1
plugin=jerasure
technique=reed_sol_van
```

Create a profile to set the data to be distributed on six OSDs (k+m=6) and sustain the loss of three OSDs (m=3) without losing data:

```
$ ceph osd erasure-code-profile set myprofile k=3 m=3
$ ceph osd erasure-code-profile get myprofile
k=3
m=3
plugin=jerasure
technique=reed_sol_van
$ ceph osd erasure-code-profile ls
default
myprofile
```

Remove a profile that is no longer in use (otherwise it will fail with EBUSY):

```
$ ceph osd erasure-code-profile ls
default
myprofile
$ ceph osd erasure-code-profile rm myprofile
$ ceph osd erasure-code-profile ls
default
```

Set the rule to ssd (instead of default):

```
$ ceph osd erasure-code-profile set myprofile \
    crush-root=ssd
$ ceph osd erasure-code-profile get myprofile
k=2
m=1
plugin=jerasure
technique=reed_sol_van
crush-root=ssd
```