the client with the location of the objects; the client must use CRUSH to *compute* the locations of objects that it needs to access.

To calculate the Placement Group ID (PG ID) for an object, the Ceph client needs the object ID and the name of the object's storage pool. The client calculates the PG ID, which is the hash of the object ID modulo the number of PGs. It then looks up the numeric ID for the pool, based on the pool's name, and prepends the pool ID to the PG ID.

The CRUSH algorithm is then used to determine which OSDs are responsible for a placement group (the *Acting Set*). The OSDs in the Acting Set that are up are in the *Up Set*. The first OSD in the Up Set is the current primary OSD for the object's placement group, and all other OSDs in the Up Set are secondary OSDs.

The Ceph client can then directly work with the primary OSD to access the object.

Data Protection

Like Ceph clients, OSD daemons use the CRUSH algorithm, but the OSD daemon uses it to compute where to store the object replicas and for rebalancing storage. In a typical write scenario, a Ceph client uses the CRUSH algorithm to compute where to store the original object. The client maps the object to a pool and placement group and then uses the CRUSH map to identify the primary OSD for the mapped placement group. When creating pools, set them as either replicated or erasure coded pools. Red Hat Ceph Storage 5 supports erasure coded pools for Ceph RBD and CephFS.

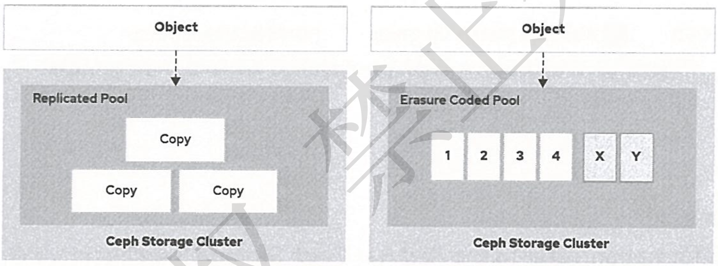

Figure 1.2: Ceph pool data protection methods

For resilience, configure pools with the number of OSDs that can fail without losing data. For a replicated pool, which is the default pool type, the number determines the number of copies of an object to create and distribute across different devices. Replicated pools provide better performance than erasure coded pools in almost all cases at the cost of a lower usable-to-raw storage ratio.

Erasure coding provides a more cost-efficient way to store data but with lower performance. For an erasure coded pool, the configuration values determine the number of coding chunks and parity blocks to create.

A primary advantage of erasure coding is its ability to offer extreme resilience and durability. You can configure the number of coding chunks (parities) to use.

The following figure illustrates how data objects are stored in a Ceph cluster. Ceph maps one or more objects in a pool to a single PG, represented by the colored boxes. Each of the PGs in this figure is replicated and stored on separate OSDs within the Ceph cluster.