CLAIMS

1. A non-transitory, computer-readable storage medium, storing program instructions that when executed by a plurality of computing devices cause the plurality of computing devices to implement:

detecting a cluster scaling event for a current cluster having a plurality of nodes storing cluster data, wherein the current cluster receives access requests for the cluster data at a network endpoint for the current cluster;

in response to detecting the cluster scaling event:

directing the current cluster to operate in read-only mode;

determining a change in the nodes of the current cluster based on the cluster scaling event;

creating a new cluster having a plurality of nodes as indicated by the change;

initiating a copy of the cluster data from the current cluster to the plurality of nodes in the new cluster; and

moving the network endpoint for the current cluster to the new cluster.