

ABSTRACT

SCALING COMPUTING CLUSTERS

A currently operating computing cluster that has multiple nodes storing cluster data may be scaled. A cluster scaling request may be received for the current cluster indicating a change in a number or type of nodes in the current cluster. In response to receiving the cluster scaling request, a new cluster may be created as indicated in the cluster scaling request, a copy of the cluster data from the current cluster to the nodes in the new cluster may be initiated, a network endpoint for the current cluster may be moved to the new cluster, and the current cluster may be disabled. The current cluster may, in some embodiments, respond to read access requests during the copy of the cluster data.

[Figure 1]