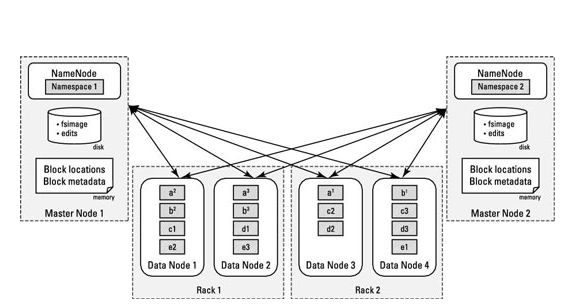
HDFS Federation:

HDFS federation improves the existing HDFS architecture through a clear separation of namespace and storage, enabling generic block storage layer. It enables support for multiple namespaces in the cluster to improve scalability and isolation.

The solution to expanding Hadoop clusters indefinitely is to federate the NameNode. Before Hadoop 2.x came into existence, Hadoop clusters has to limitation on the degree to which they could scale. Few clusters were able to scale beyond 3000 or 4000 nodes.

NameNode's need to maintain records for every block of data stored in the cluster turned out to be the most significant factor restricting greater cluster growth.

When you have two many blocks, it becomes increasingly difficult for the NameNode to scale up as the Hadoop cluster scales out.



HDFS Federation block diagram.

Specially you must set HDFS up so that you have multiple NameNode instances running on their own, dedicated master nodes and then making each NameNode responsible only for the file blocks in its own name space.