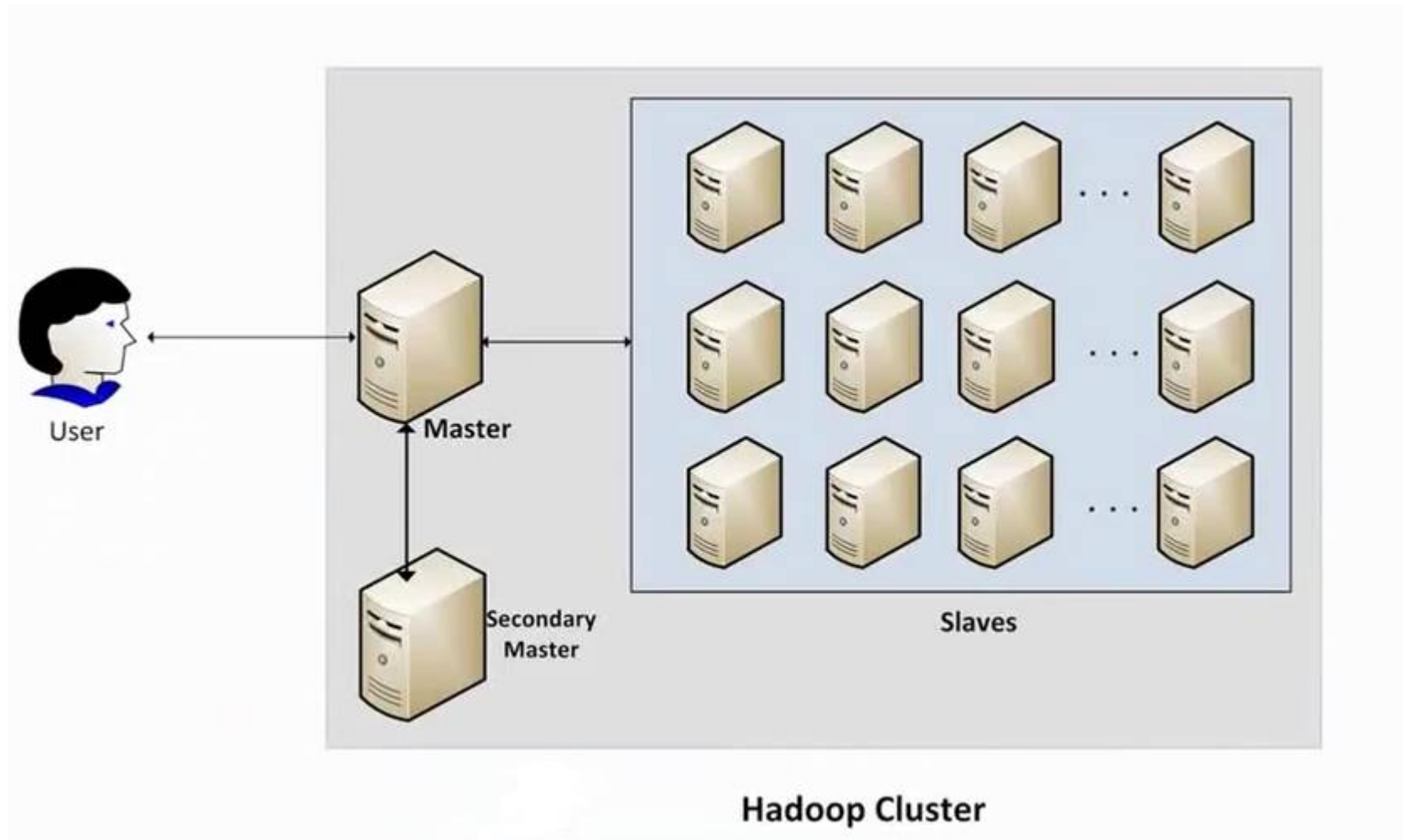# Hadoop Basic Architecture
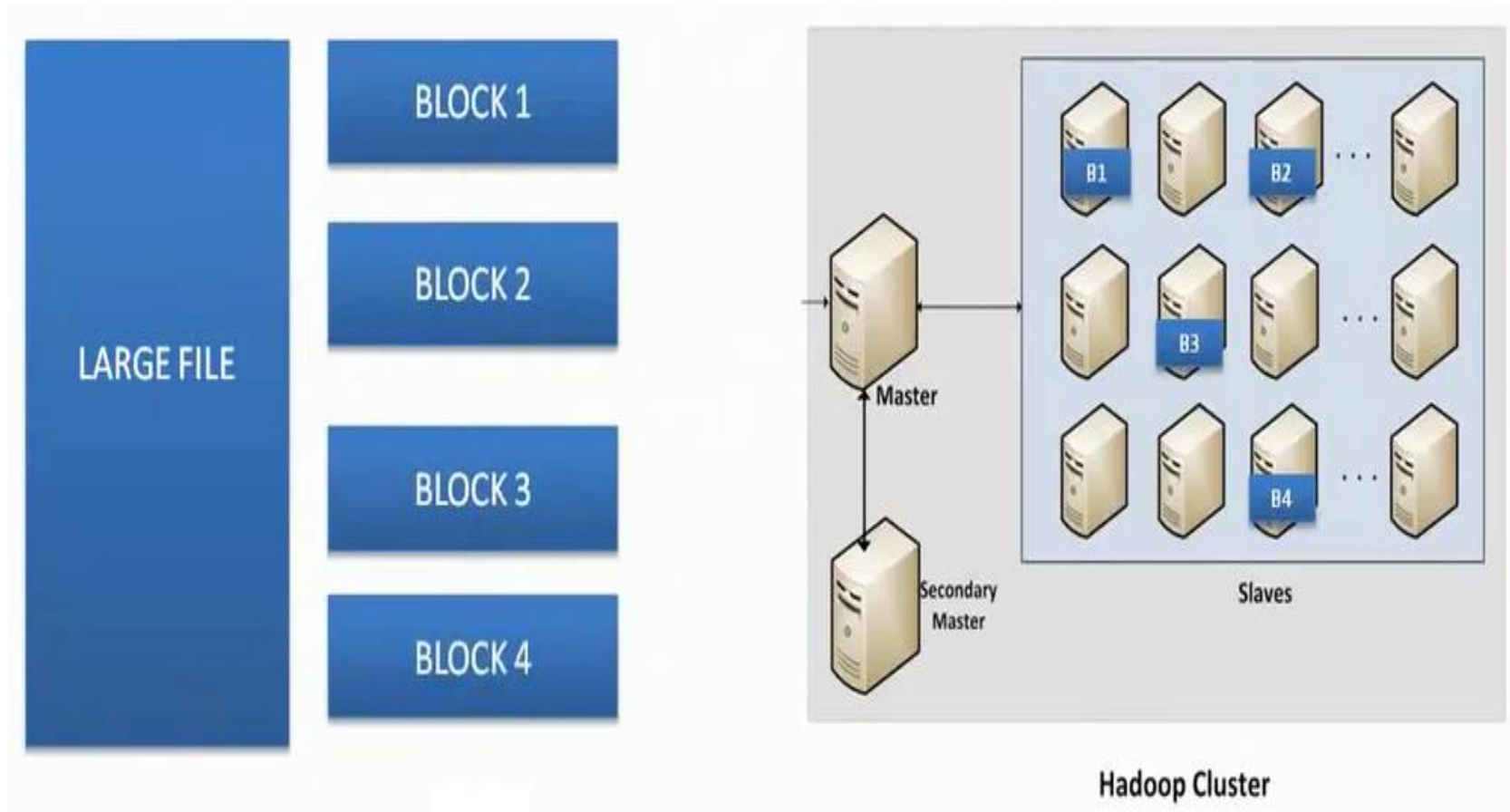
HDFS Characteristics

✓ HDFS is a highly fault tolerant, distributed, reliable, scalable file system for data storage.

✓ HDFS stores multiple copies of data on different nodes; a file is split up into blocks (default 64 MB) and stored across multiple machines (by default 3 machines)

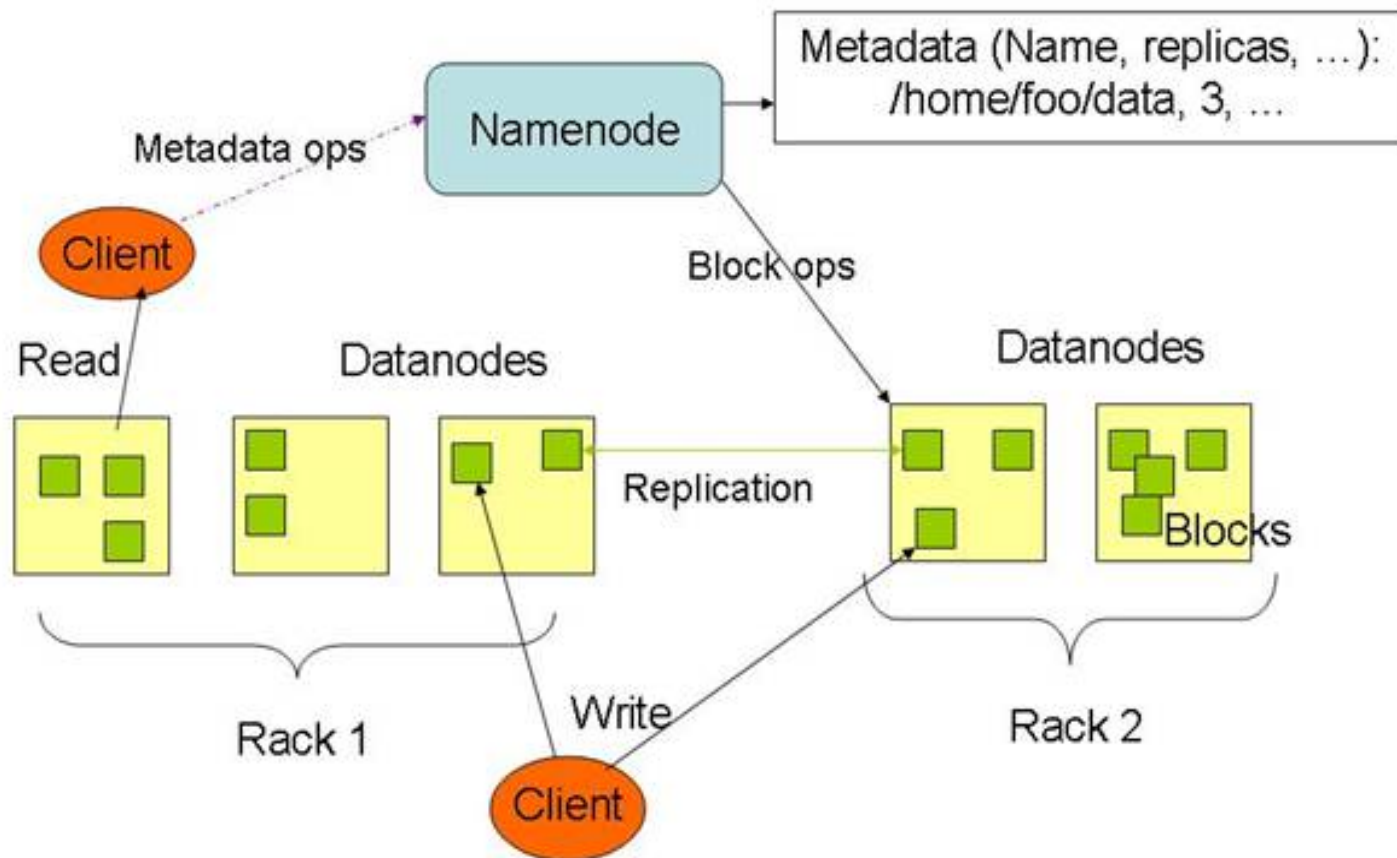✓ Hadoop cluster typically has a single namenode and number of datanodes to form the HDFS cluster

HDFS architecture

HDFS Architecture

Map Reduce

✓ Map-Reduce is a programming model designed for processing large volumes of data in parallel by dividing the work into a set of independent tasks

✓ Map-Reduce is a paradigm for distributed processing of large data set over a cluster

Hadoop Cluster