

HDFS

The Hadoop Distributed File System

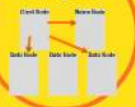
Overview



HDFS Architecture



Reading a File



Writing a File



Secondary Node



Using HDFS

1. Upload
2. Download
3. Delete
4. Rename
5. Move

Let's
Play

HDFS

The Hadoop Distributed File System

Overview

Handles big files



By breaking them into blocks



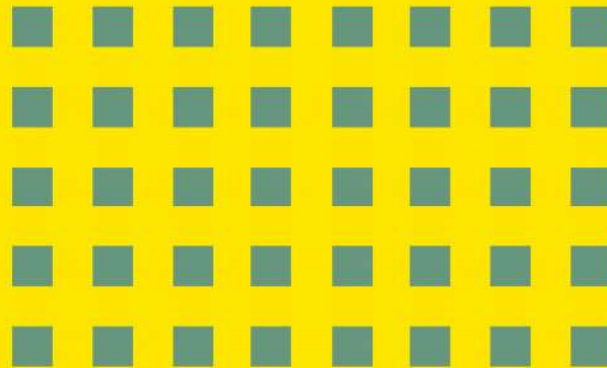
Stored across several
commodity computers



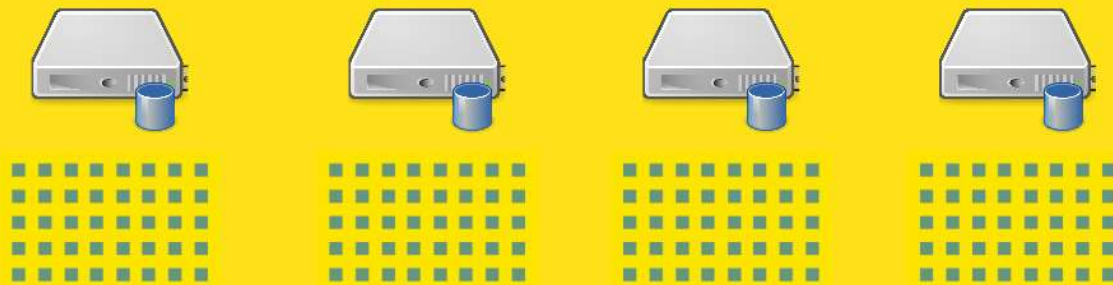
Handles big files



By breaking them into blocks



**Stored across several
commodity computers**



HDFS Architecture

Name Node



Data Node



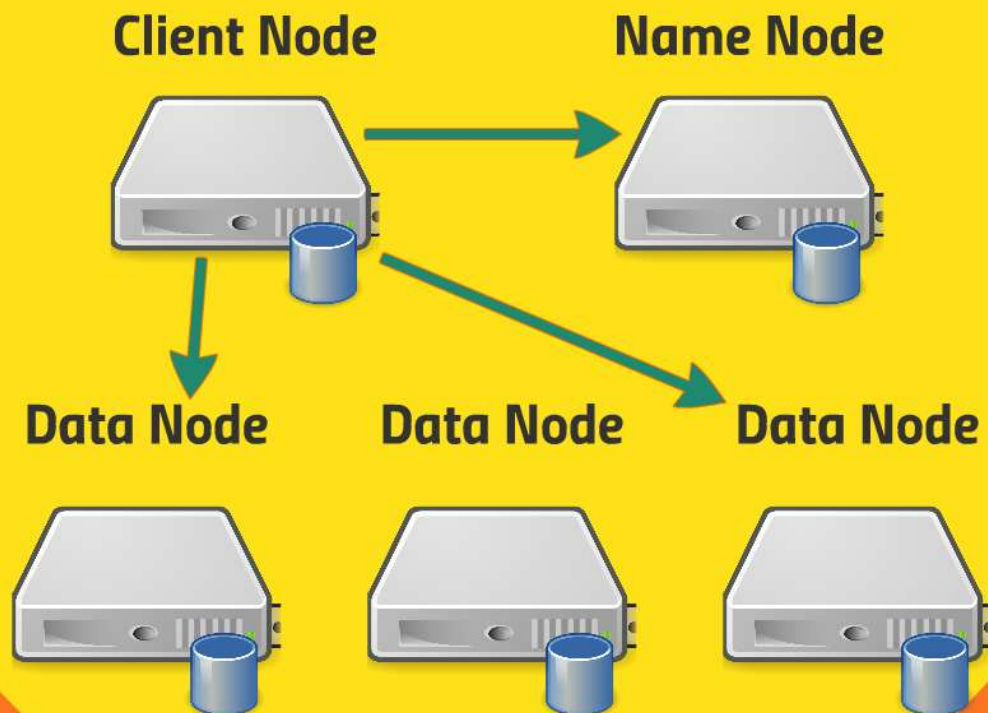
Data Node



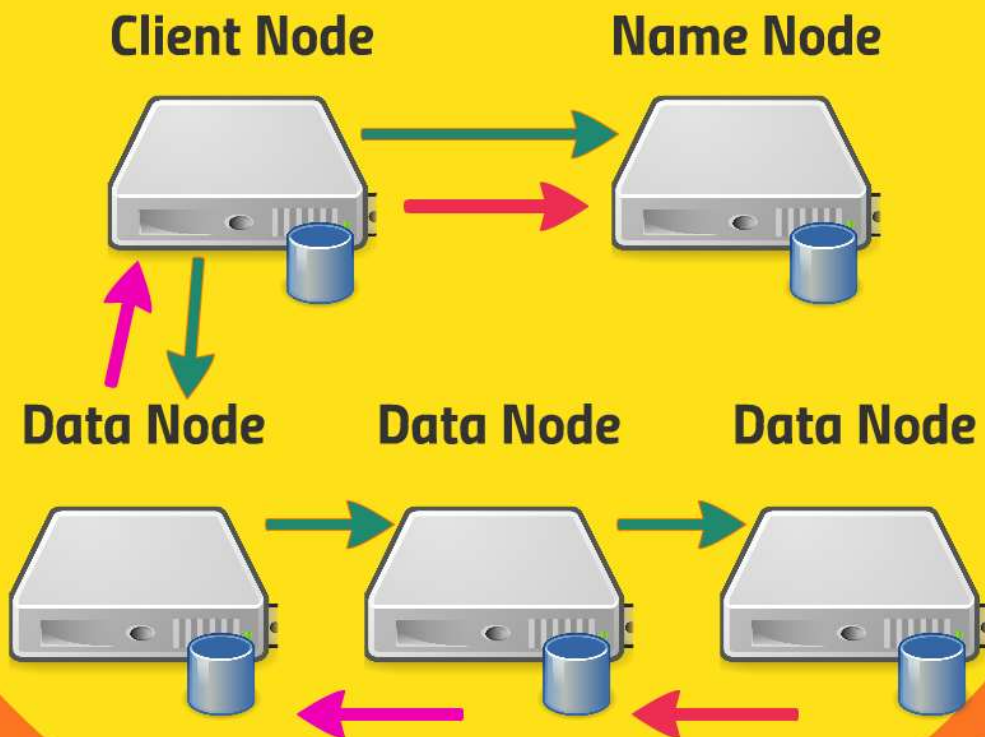
Data Node



Reading a File



Writing a File





Back Up Metadata

**Namenode writes to local
disk and NFS**

Secondary Namenode

**Maintains merged copy
of edit log you can
restore from**

Using HDFS

UI (Ambari)

Command-Line Interface

HTTP / HDFS Proxies

Java interface

NFS Gateway