

HDFS

The Hadoop Distributed File System

Overview



HDFS Architecture



Reading a File



Writing a File



Hardware Architecture



Using HDFS

1. Check the Hadoop installation and configuration.
2. Set up the HDFS environment.
3. Use the HDFS commands to manage the data.

Let's
Play

HDFS

The Hadoop Distributed File System

Overview

Handles big files



By breaking them into blocks



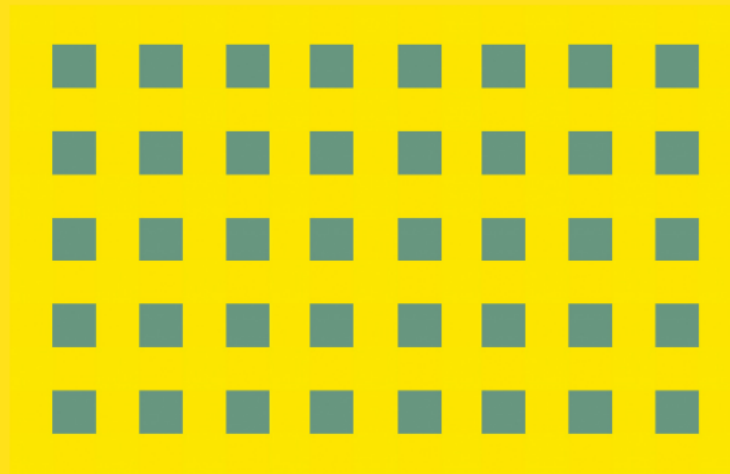
Stored across several
commodity computers



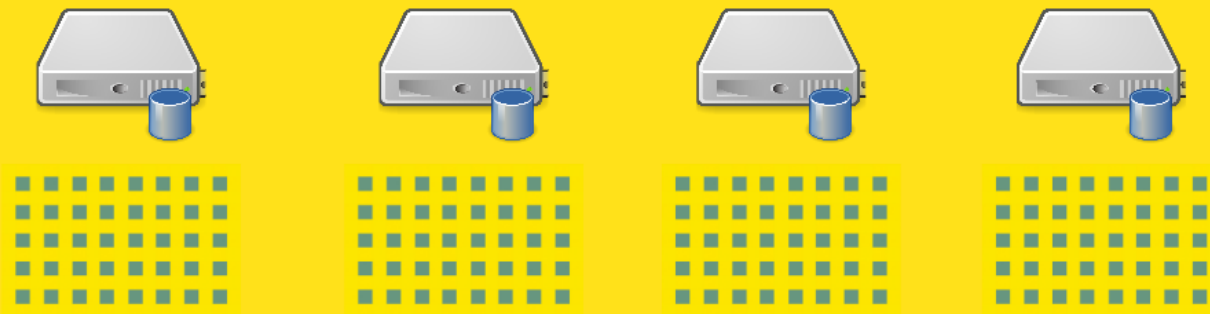
Handles big files



By breaking them into blocks



**Stored across several
commodity computers**



HDFS Architecture

Name Node



Data Node



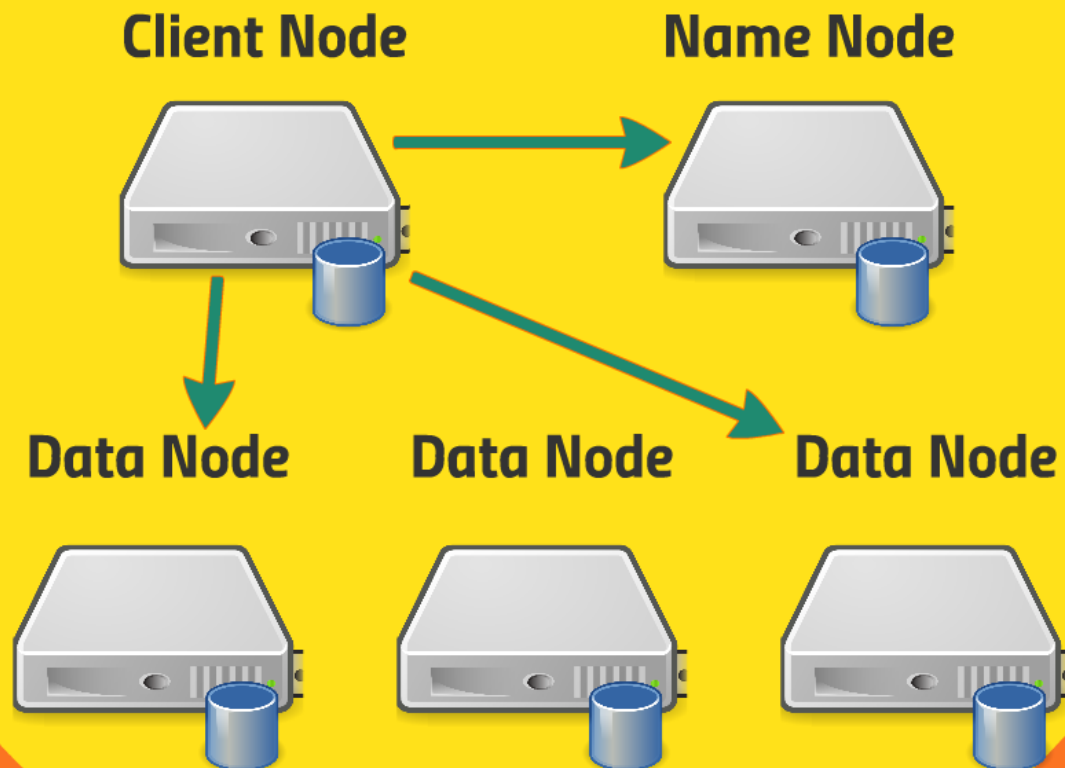
Data Node



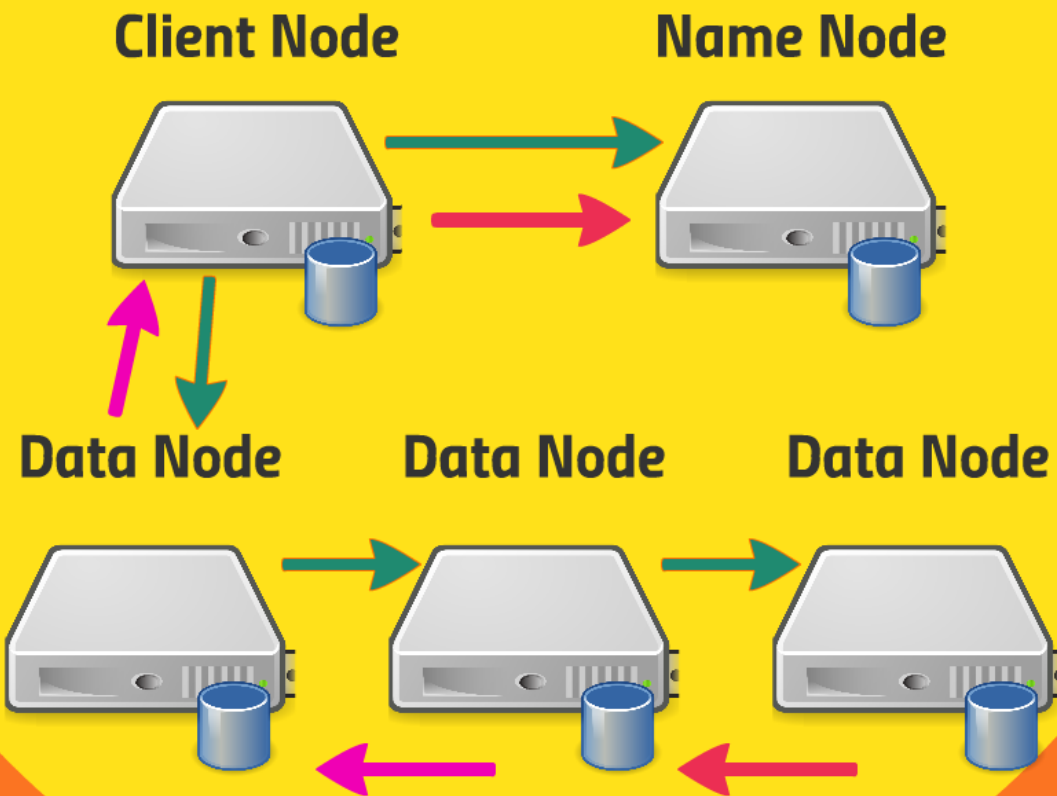
Data Node



Reading a File



Writing a File





Namenode Resilience

Back Up Metadata

Namenode writes to local disk and NFS

Secondary Namenode

Maintains merged copy of edit log you can restore from

HDFS Federation

Each namenode manages a specific namespace volume

HDFS High Availability

- Hot standby namenode using shared edit log
- Zookeeper tracks active namenode
- Uses extreme measures to ensure only one namenode is used at a time

Back Up Metadata

**Namenode writes to local
disk and NFS**

Secondary Namenode

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

HDFS Federation

**Each namenode
manages a specific
namespace volume**

HDFS High Availability

- **Hot standby namenode using shared edit log**
- **Zookeeper tracks active namenode**
- **Uses extreme measures to ensure only one namenode is used at a time**

Using HDFS

UI (Ambari)

Command-Line Interface

HTTP / HDFS Proxies

Java interface

NFS Gateway



**Let's
Play**