Distributed File System

```
┌─────────────────────┐
│  Name Node          │
└─────────────────────┘
   │        │        └──────┐
┌──────┐  ┌──────┐      ┌──────┐
│ Data │  │ Data │      │ Data │
│Node 1│  │Node 2│      │Node n│
└──────┘  └──────┘      └──────┘
```

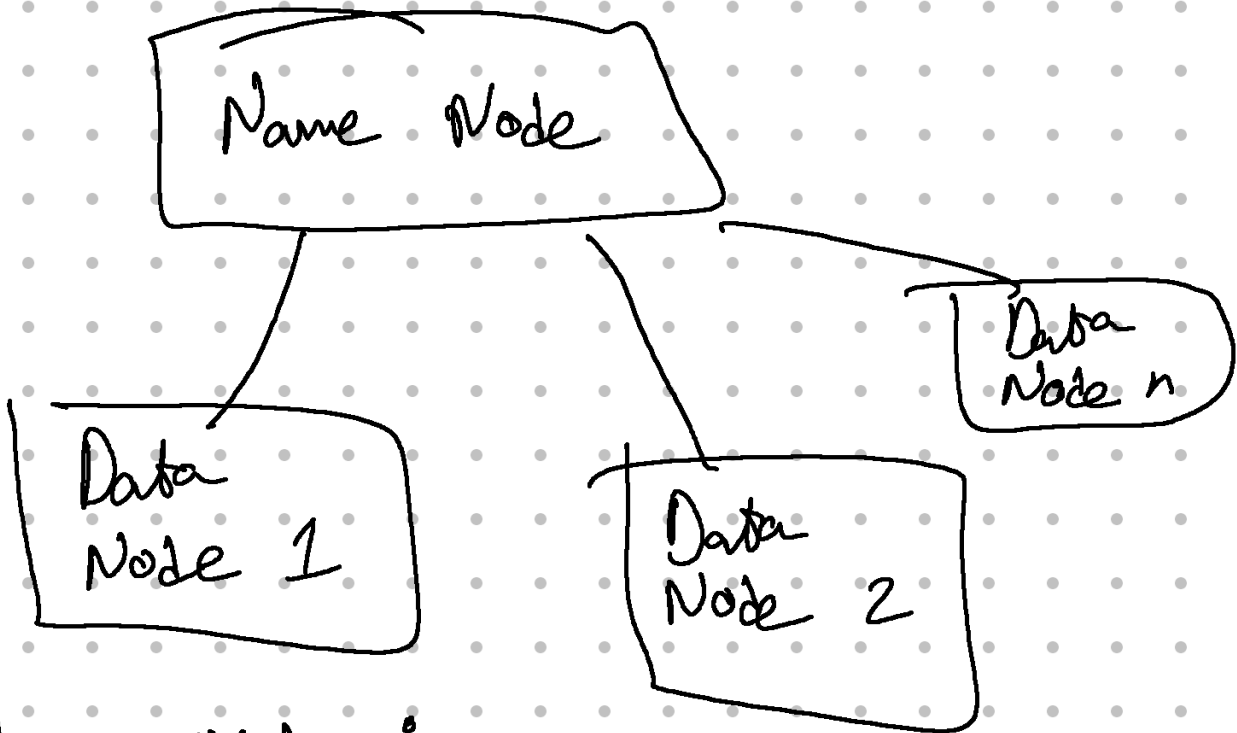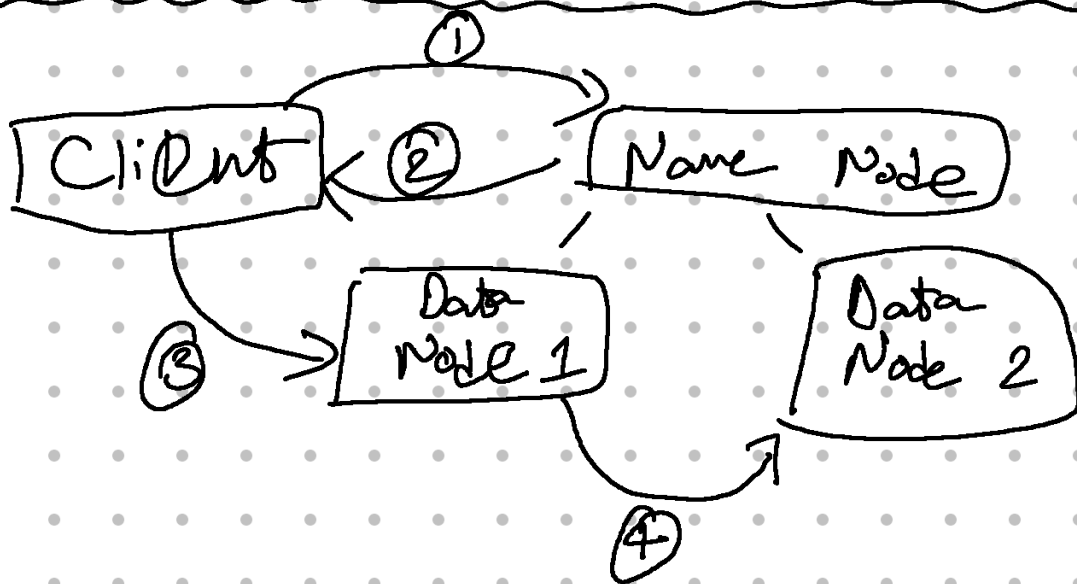## Name Node :

• Stores metadata of the file system

• is aware of where the chunks of a file are stored

• acts like a master node.

• is aware of data node size

• does health check of data nodes
  ↳ takes care of machines that fail)
  and replicates data as necessary

• Can be aware of data centre geography
  ↳ for eg. which rack a data node is in
  useful for when you don't want to
  replicate a chunk in the same rack

⇒ Block size determines how small each block will be broken down into.

⇒ Replication Factor...

---

① Client requests to store file at name node

② Data node instructs client which node to store file

③ Client stores at Data Node 1

④ Data Node 1, replicates data on data node 2 upon instruction from name node