

Why Big Data now?

- More data are being collected and stored
- Open source code
- Commodity hardware / Cloud

-
- High-Volume
 - High-Velocity
 - High-Variety

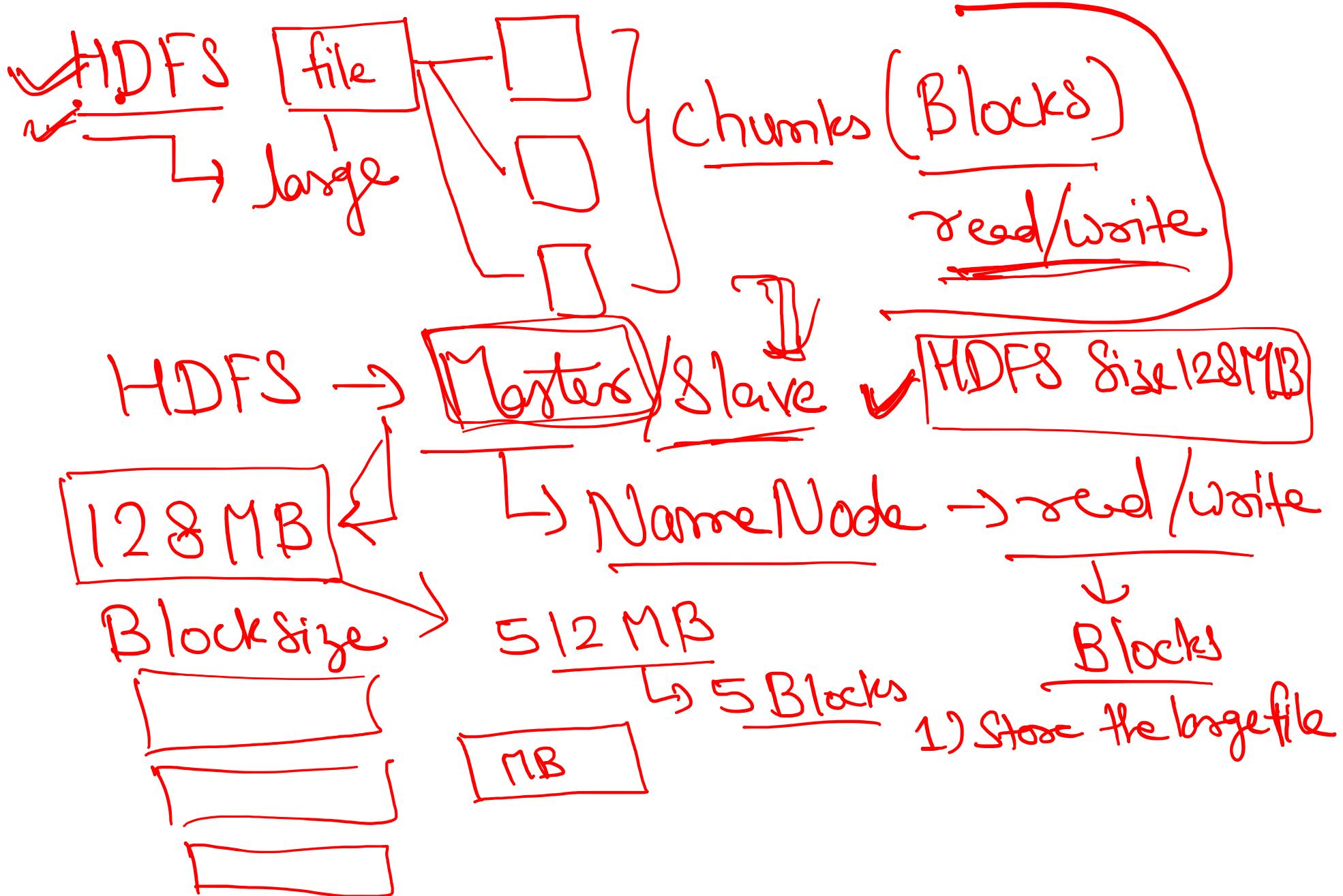
TV's
Value is big data.

→ Artificial
Intelligence

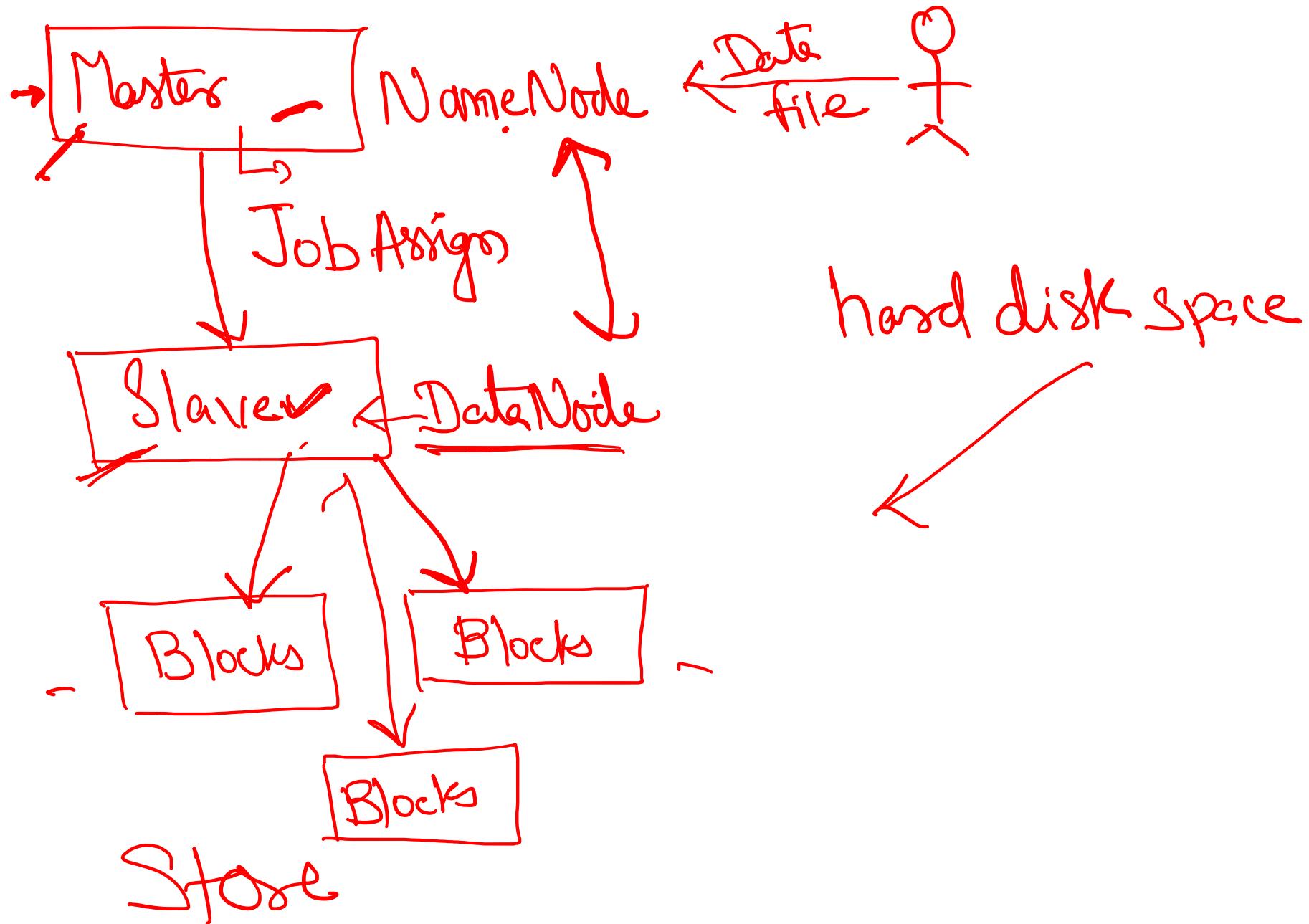
Hadoop Blocks – Namenode

~~Date~~

Blocks ✓



Hadoop Blocks – Datanode ✓

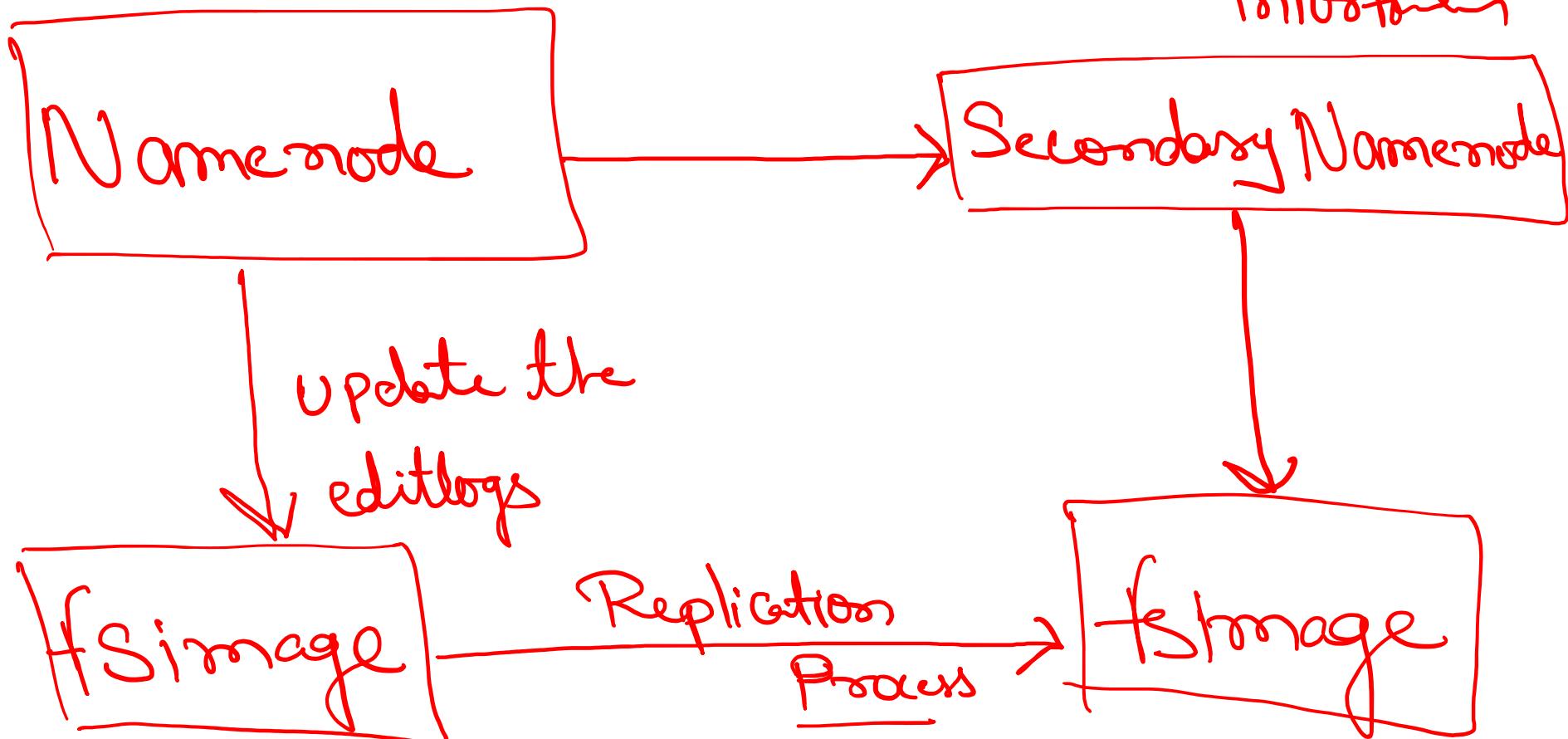


Hadoop Blocks – Secondary Namenode

fstorage- Stores the snapshot of its

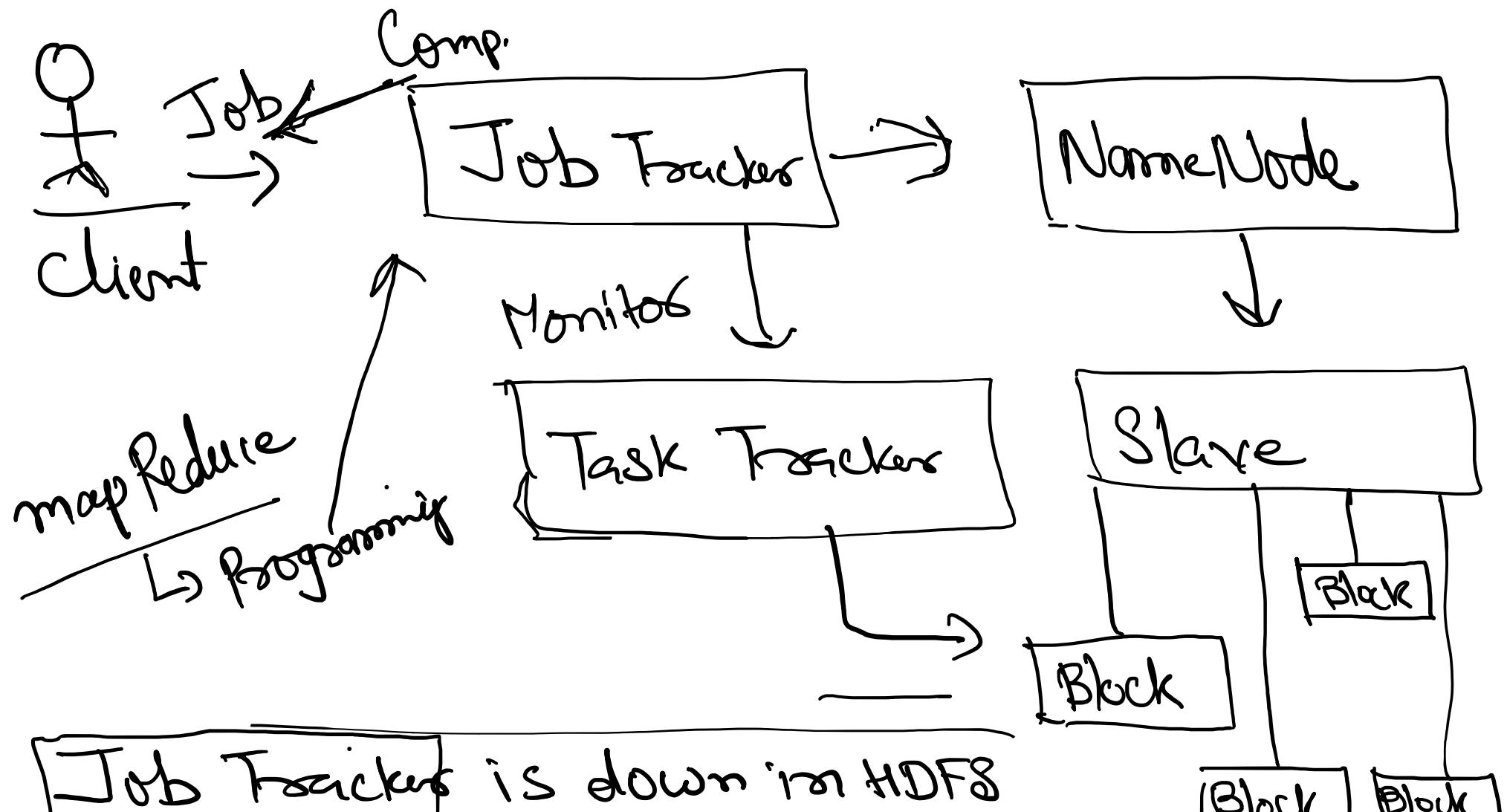
Namenode - holds metadata for HDFS in blocks

infoform



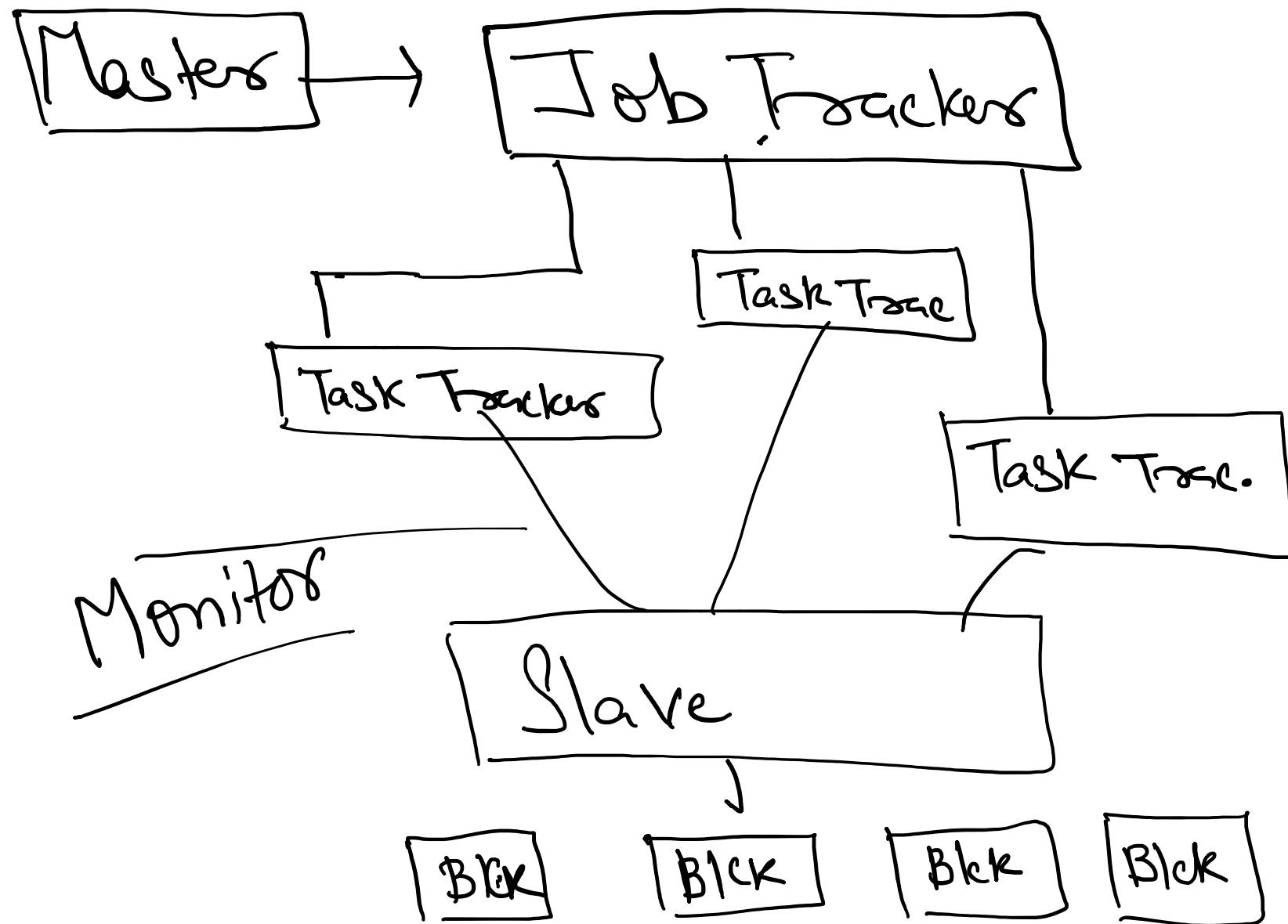
Editlogs → keeps track of every change of HDFS.

Hadoop Blocks – Job Tracker ✓



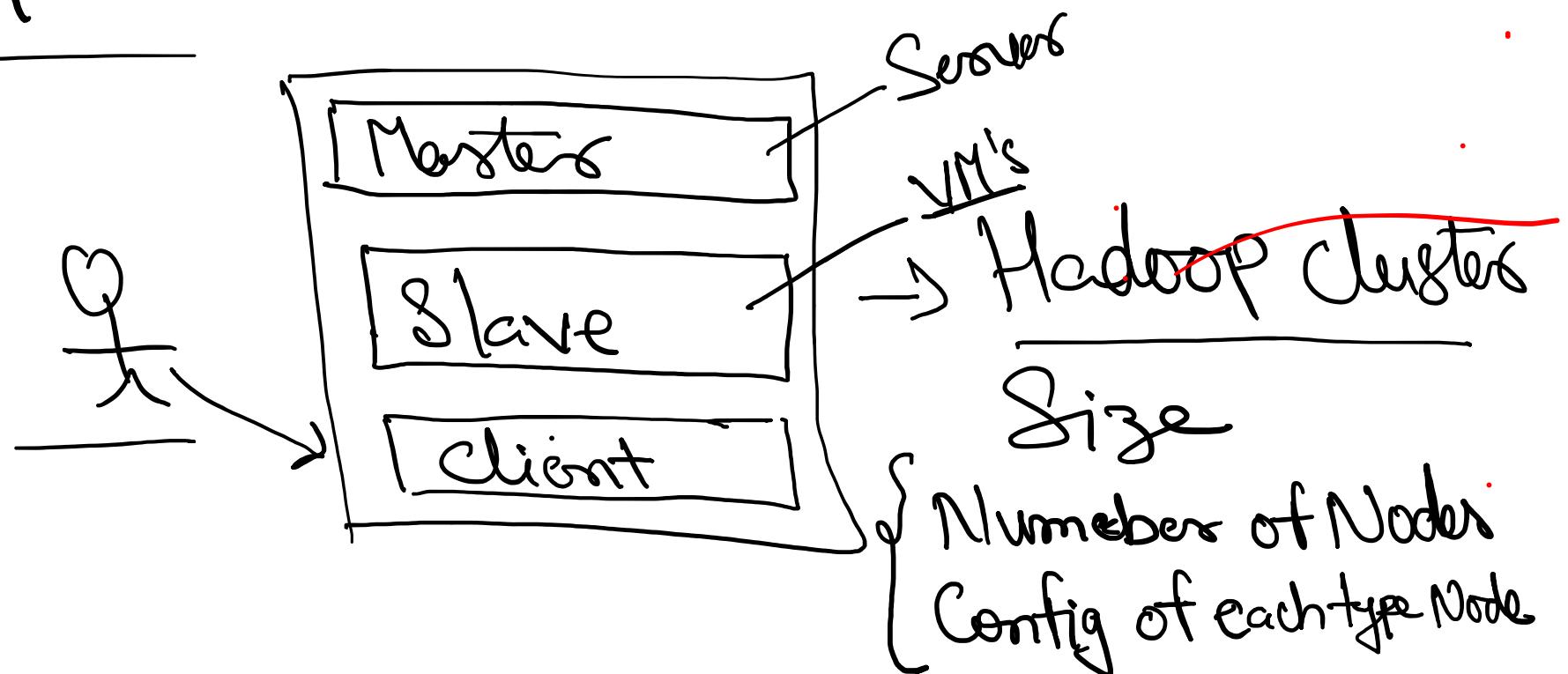
mapReduce → halt

Hadoop Blocks – Task Tracker



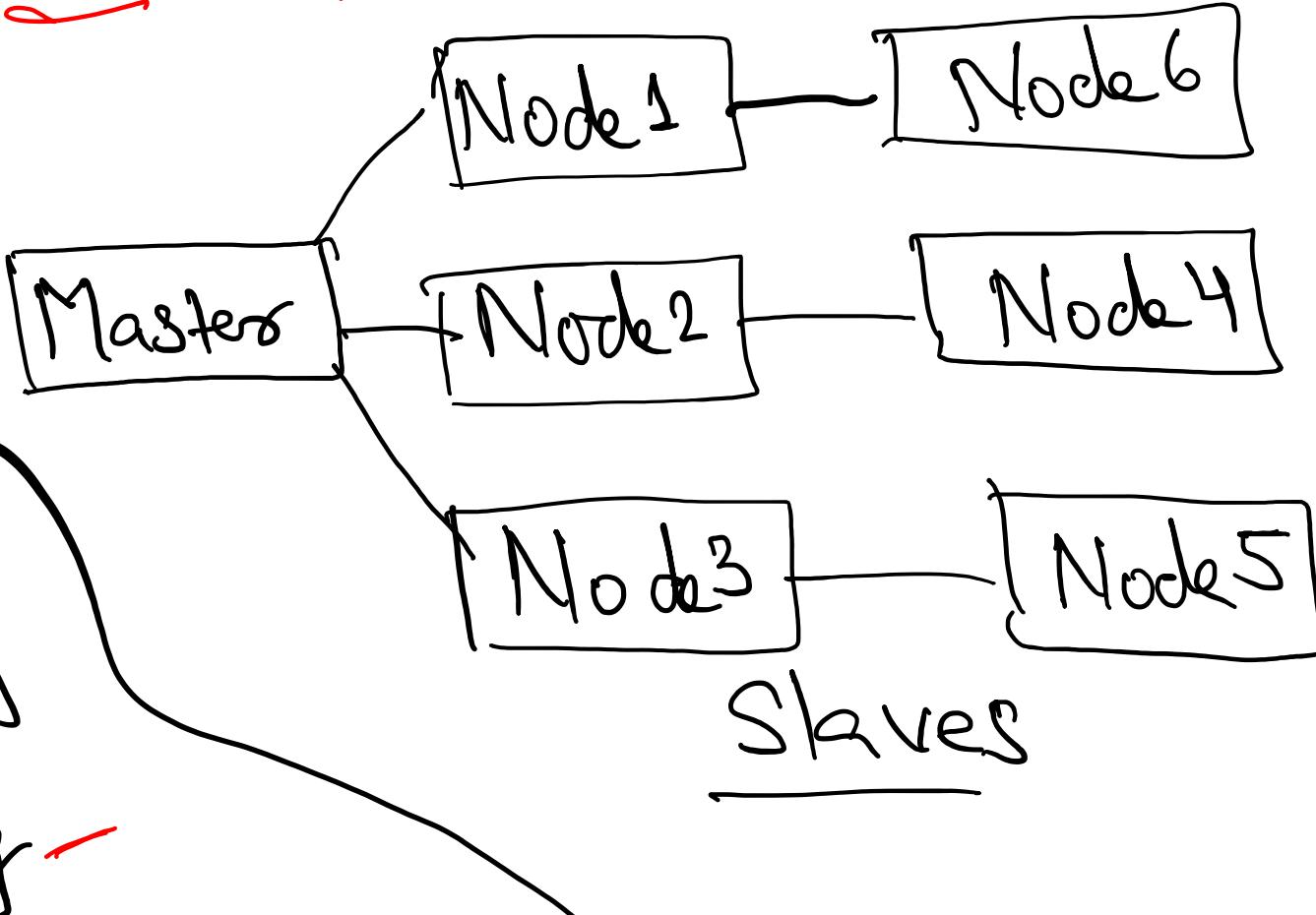
Hadoop Cluster - Introduction ✓

Apache ⇒ big data analytics, process tasks to be broken down into smaller tasks that can be performed in parallel by using any algo like mapReduce.



Data Store — Struct, Semi-Struct & Unstruct Data

Schemas =>



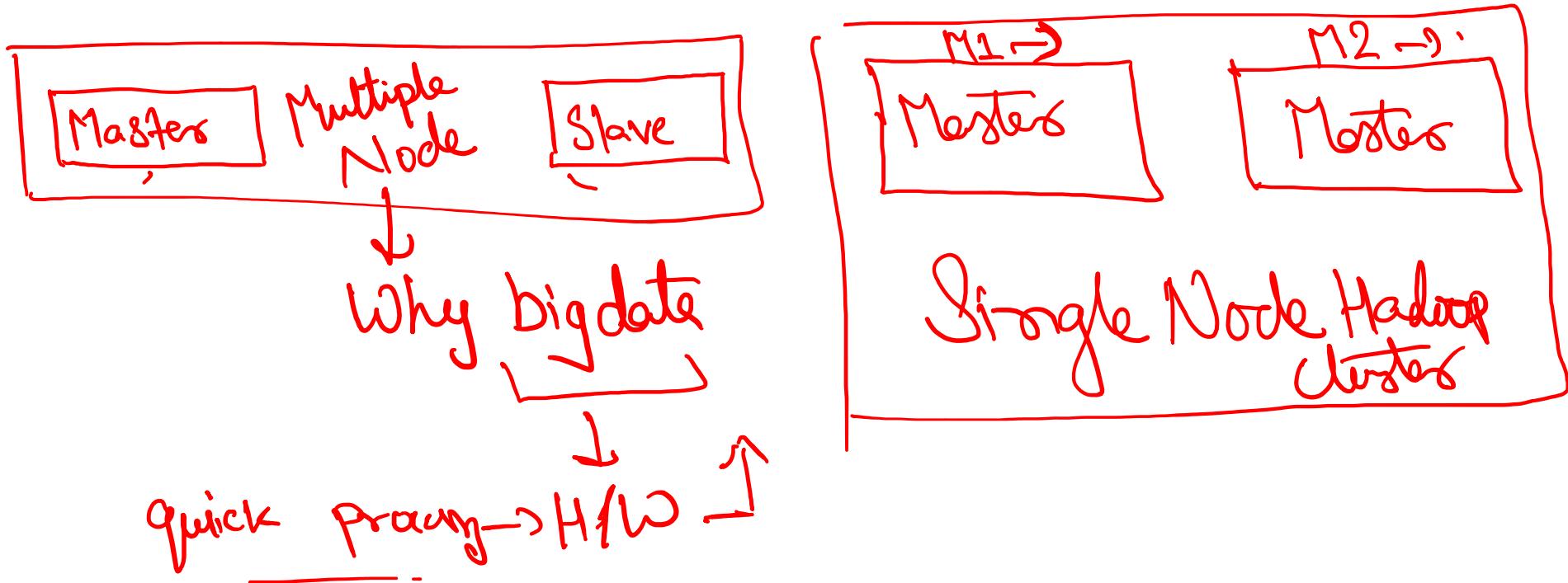
Properties ✓

- Scalability
- flexibility
- Speed
- No data loss
- Economical

Slaves

Type of Hadoop Cluster

1) Single Node Hadoop | 2) Multiple Node



Types of Mode of Operation

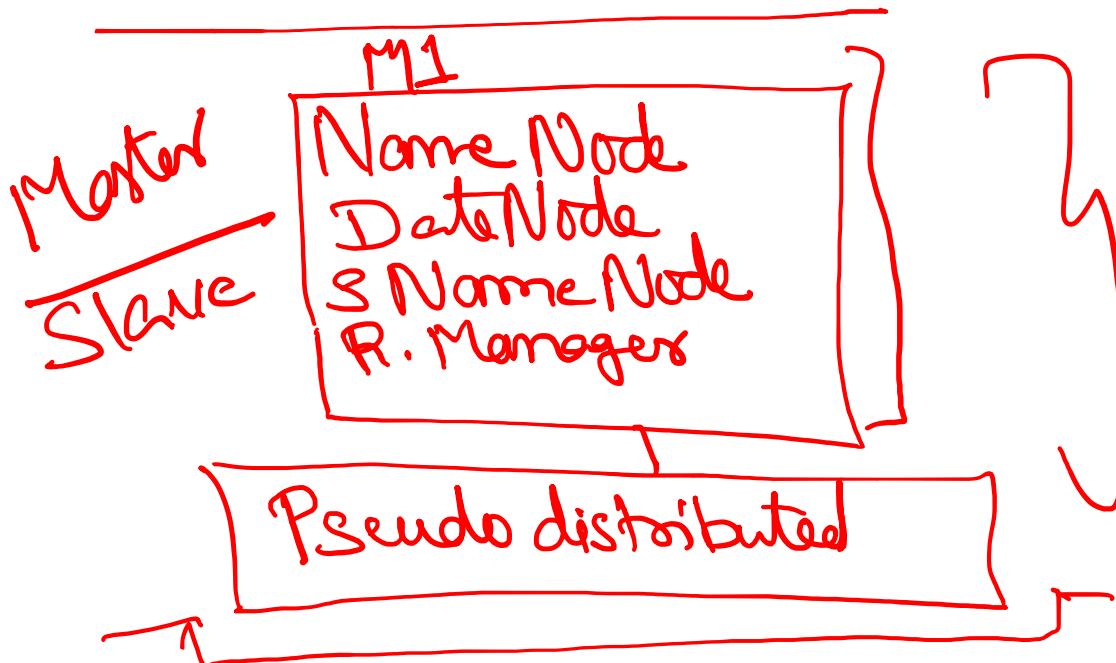
1). Standalone Mode → M1 (All nodes execute)



Fastest

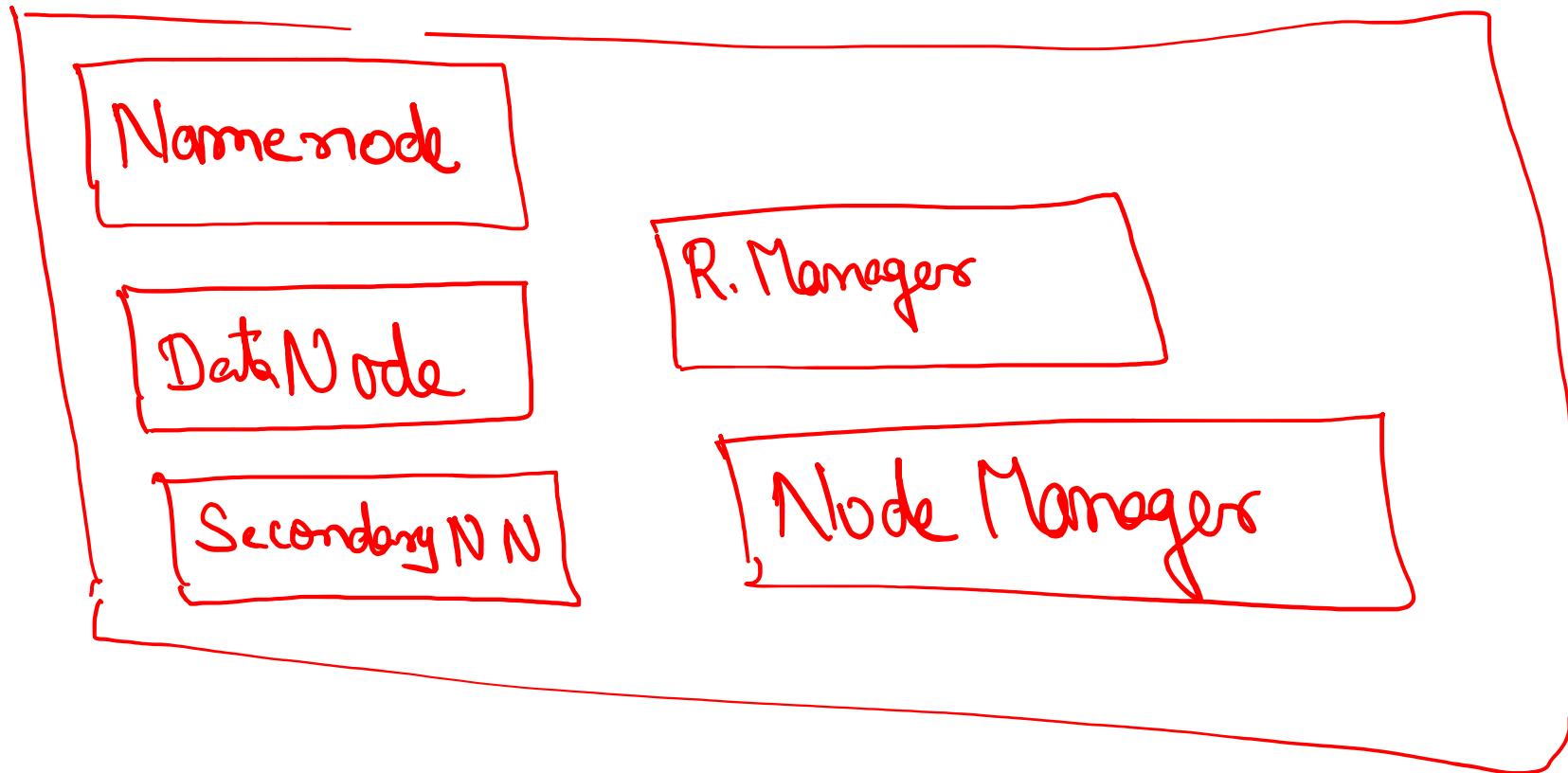
M2 (Manage & Monitoring purpose)

2) Pseudo Distributed Mode → ✓

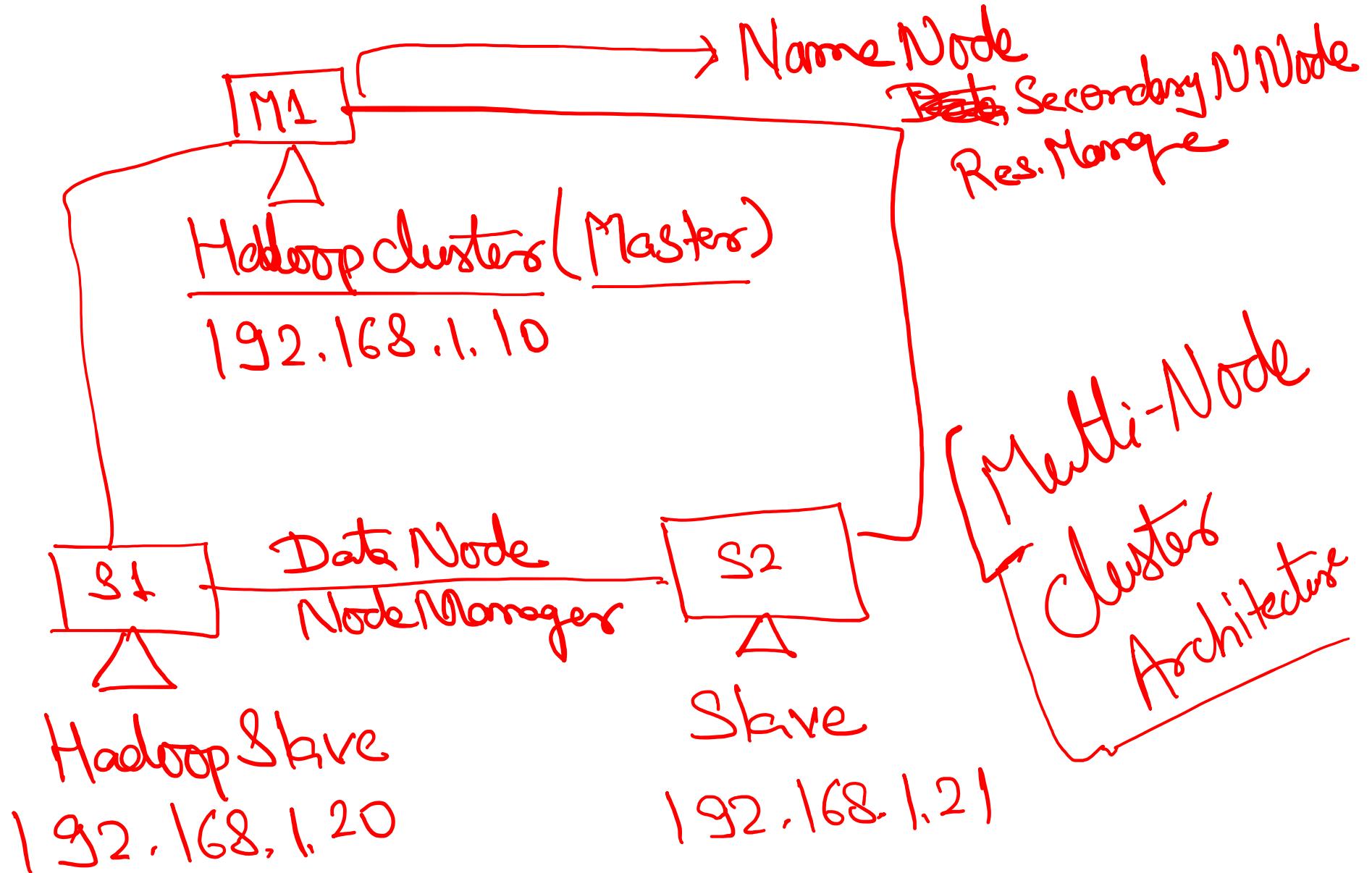


Mode env. is used
in development and
debugging process.

3). Fully Distributed – Multi Node clusters(↗)



fully Distributed Mode



Config file - [Imp] → Apache Hadoop docs.
{URB3}

1). HADOOP-ENV.sh → JDK

[\$JAYA_Home]

2) CORE-SITE.XML → Cluster Info

↳ NameNode - IP, Port#,

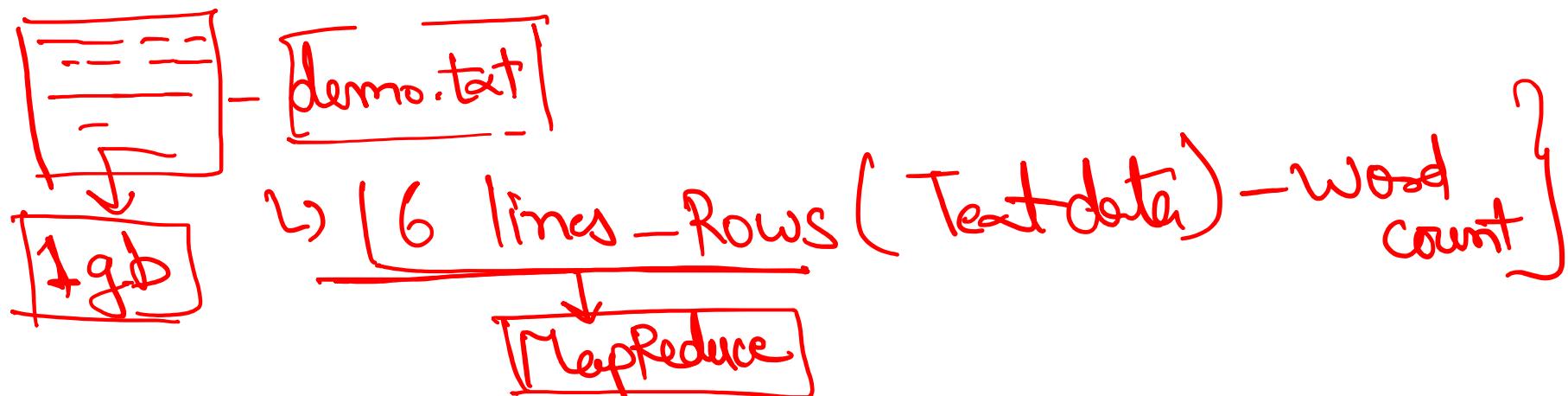
3) HDFS-SITE.XML → Replication of Block

4) MAPRED-SITE.XML → MapReduce Program
Setting

5) Masters -

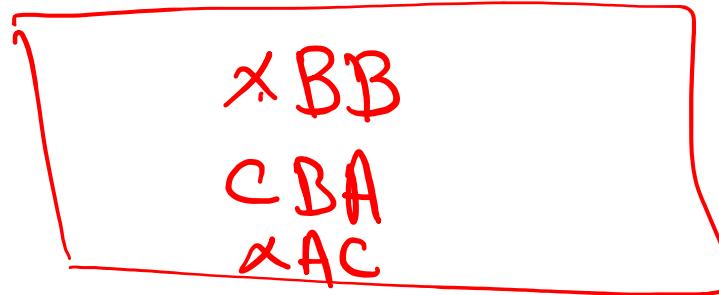
6) Slave - Masters Node info a list of hosts, one per line
IP address of Slave nodes.

MapReduce \Rightarrow it is responsible for processing the
data file in hadoop ecosystem



Map Reduce

- Map Task
- Reduce Task



XBB

Example of MapReduce

Program →

Input

file

X B B
C B A
X A C

SPLIT

X BB

CBA

X AC

MAP

X B B
C B A
X A C

C B A
X A C

X A C

COMBINE

A 1
A 1

B 1
B 1
B 1

C 1
C 1

X 1
X 1

Reduce

A 2
B 3
C 2
X 2