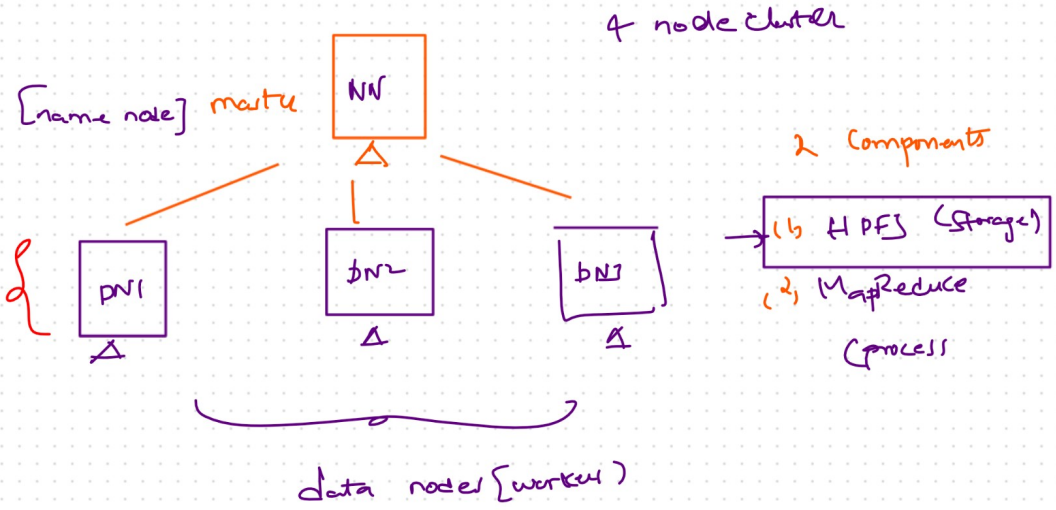
Node failures in Hadoop

* Hadoop uses master-slave architecture



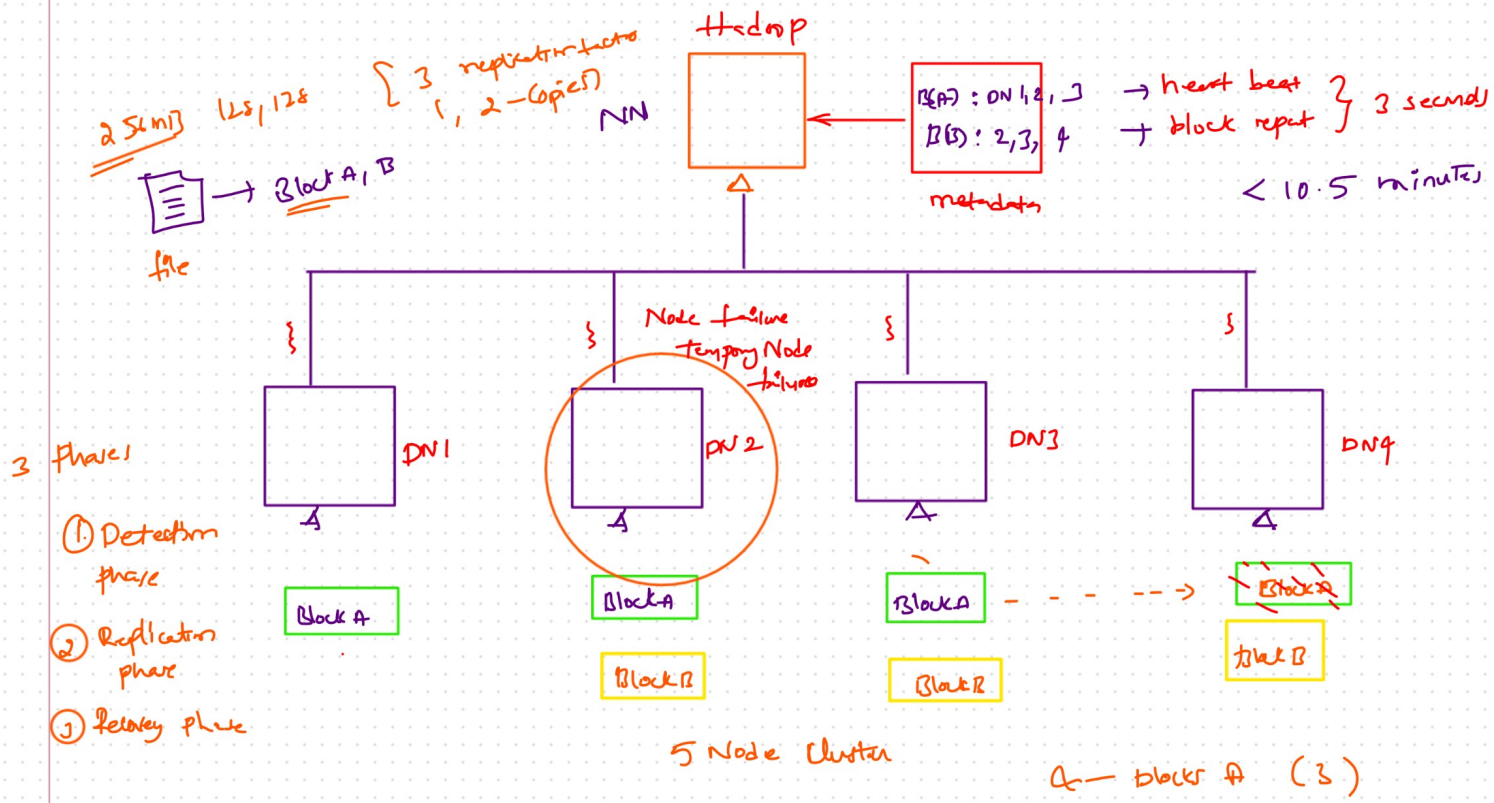
* replication & rack awareness helps us to safeguard against data loss ✓

Types of Data Node failures

* Temporary Node failures
* Permanent Node failures

Temporary Data Node failures

* Network outages
* Software Crashes (OS or applications)
* Node reboots for maintenance
* Routine patch installations



① Data Node 2 fail — no heartbeat for 10.5 minutes

② Name node marks block on Data 2 as under-replication

③ Name node will trigger replication

— Copy Block A to Data Node 4

→ Now Block A exists in 1, 3, 4 (Data Nodes)

④ Data Node 2 recovers & sends block report

⑤ Name node detects excess replica of Block A

i.e present in 4 data Nodes (1, 2, 3 4)

⑥ Name node deletes excess replica from one node

⑦ Metadata updated to maintain Consistency

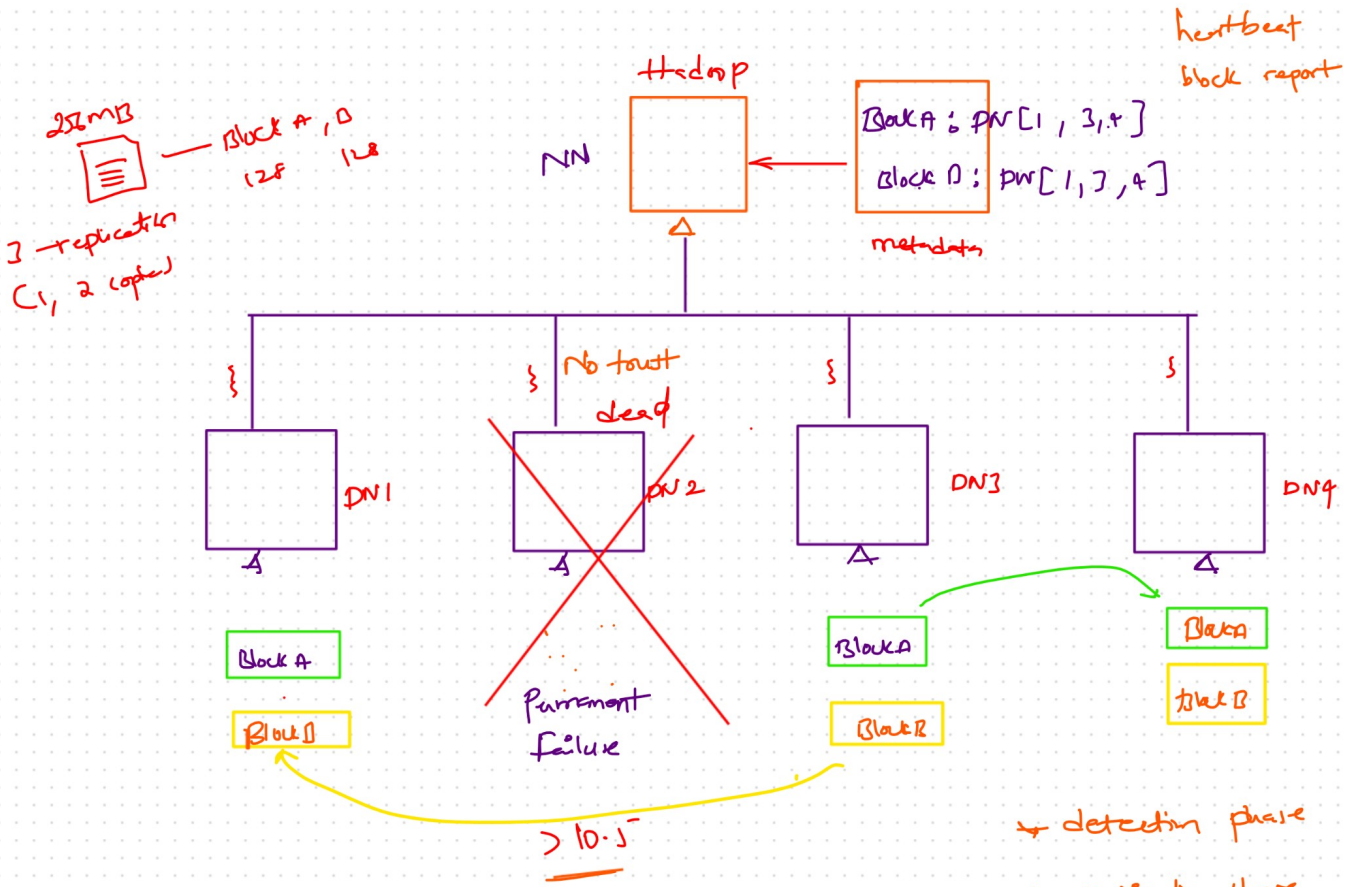
Key Points:

* No operations are affected during temporary failures
* Metadata is always kept consistent
* The system automatically maintains the configured replication factor
* The entire process is handled by Hadoop in the background

Permanent Data Node failure :

Causes :

+ hardware failure
+ disk corruption
+ decommissioning (intentionally remove a data node from cluster)



+ detection phase
+ replication phase
+ Data Node return behavior

Detection Phase :

(1) Name node identifies a data node is dead after missing heartbeats over an extended period (> 10·J).

Replication Phase :

(1) Blocks stored on failed data nodes are considered permanently lost

(2) Name Node initiates replication process to restore data

3) Specific blocks from the failed data node are identified & replicated to healthy nodes

4) Metadata is updated to reflect new replicas

5) References to failed data node are moved from metadata

Data Node Return Behaviour:

1) if failed node returns after prolonged downtime (≥ 10.5), it is not trusted

2) Blocks on recovery node are marked as stale

3) The data node is considered as fresh node

4) Existing blocks on recovered node are ignored & cleaned up to avoid data inconsistency

Diff b/w Temporary & Permanent failures

Aspect	Temporary failure	Permanent failure
Detection	missing heartbeats (< 10.5 minutes)	missing heartbeats (> 10.5 minutes)
Replication	may initiate but expects node return / recovery	Full replication to maintain factor
Node Recovery	Trusted & reconciles blocks	Treated as new node, no trust & blocks are marked stale
Meta data	Dynamic updates	Permanent removal of node references